



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE INFORMÁTICA  
GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

# Modelo de Transferência de Aprendizagem baseado em Regressão Linear Regularizada

**Larissa Camila Ferreira da Silva**

RECIFE, 22 DE JUNHO DE 2017

LARISSA CAMILA FERREIRA DA SILVA

MODELO DE TRANSFERÊNCIA DE APRENDIZAGEM BASEADO EM REGRESSÃO  
LINEAR REGULARIZADA

Monografia apresentada ao curso de Engenharia de Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Ricardo Bastos C. Prudêncio.

RECIFE  
2017

LARISSA CAMILA FERREIRA DA SILVA

MODELO DE TRANSFERÊNCIA DE APRENDIZAGEM BASEADO EM REGRESSÃO  
LINEAR REGULARIZADA

Monografia apresentada ao curso de Engenharia de Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Ricardo Bastos C. Prudêncio.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Ricardo Bastos C. Prudêncio (Orientador)  
Universidade Federal de Pernambuco (UFPE)

---

Prof. Dr. Cleber Zanchettin  
Universidade Federal de Pernambuco (UFPE)

## **AGRADECIMENTOS**

Gostaria de agradecer primeiramente a Deus, que esteve comigo durante todos os dias da minha vida, e sem ele eu não teria chegado até aqui. Sou verdadeiramente grata, pois ele foi meu fiel amigo, refúgio e fortaleza.

A minha família, principalmente a meus pais que me ajudaram a chegar aqui, me educaram e guiaram no caminho certo. Também aos meus irmãos, Luiz, Danilo e Mayara que me apoiaram de uma forma maravilhosa durante todos esses anos.

Aos meus queridos irmãos da Igreja Cristã Maranata do Rosarinho que me apoiaram, torceram e oraram por mim, e me mostraram como é viver em um corpo unido.

Ao professor Ricardo Prudêncio, que foi meu orientador e me deu todo o apoio necessário para a realização deste trabalho de graduação. Ao professor M. Brian Blake (EUA) que me proporcionou o primeiro contato com a área de Inteligência Artificial, a qual eu pretendo seguir estudando em uma pós-graduação.

Aos meus grandes amigos e companheiros de faculdade Marina Bezerra, Maria Luiza, Jilmar Teixeira, Leonardo da Vince, Caio Monte, Allysson Lukas e Luiz Henrique, por todo companheirismo durante várias disciplinas cursadas ao longo da graduação. Em especial para meu amigo Laerte da UFC, que foi muito importante para mim durante meus estágios nos EUA, me ajudando nas inúmeras pesquisas científicas, sendo um excelente profissional e amigo.

“O Senhor Deus é a minha força, e fará os meus pés como os das cervas, e me fará andar sobre as minhas alturas.”

Habacuque 3:19, Bíblia

## RESUMO

Muitos algoritmos da área de Aprendizagem de Máquina precisam de um bom conjunto de dados de treinamento para gerarem bons resultados nos dados de teste, onde estes dois tipos de dados devem ser de um mesmo contexto. Entretanto, em muitos problemas do mundo real não há dados suficientes para treinamento, e nestes casos se faz necessário o uso de técnicas de Transferência de Aprendizagem de um contexto para outro. Este trabalho apresenta uma técnica para modelagem de mudança de contexto em problemas de regressão linear. Neste método, a escolha de um atributo determina a divisão de um contexto em dois. O contexto que possui um conjunto maior de dados de treinamento é denominado contexto de origem, tem maior distribuição, e para ele um modelo é aprendido. O segundo é chamado contexto de destino, e seus dados são usados como teste no modelo do primeiro. Para que haja uma adaptação entre os ambientes, uma função de erro de aproximação é gerada através de uma regressão linear regularizada com validação cruzada. Resultados experimentais com base em uma série de testes com dados reais são mostrados.

## **ABSTRACT**

Many algorithms in the Machine Learning area have the need for a good training dataset to generate good results in the test data, where those two types of data must be in the same context. However, in many real-world problems, there is insufficient data for training, and in these cases, it is necessary to use Transfer Learning techniques from one context to another. This paper presents a technique for modeling context change on linear regression problems. In this method, the choice of an attribute determines the division of a context into two. The context that has a larger set of training data is called the source context, it has a larger distribution, and a regression model is learned for it. The second context is called target context, and its data is used for testing in the first model learned. For an adaptation among the environments, an approximation error function is generated through a regularized linear regression through cross validation. Experimental results based on a series of tests with real data are shown in this paper.

## SUMÁRIO

1. INTRODUÇÃO	9
1.1. Motivação	9
1.2. Objetivos	9
1.3. Estudo de Caso	9
1.4. Organização	10
2. FUNDAMENTAÇÃO TEÓRICA	11
2.1. Aprendizagem de Máquina	11
2.1.1. Aprendizagem supervisionada	11
2.1.2. Aprendizagem não supervisionada	13
2.2. Regressão Linear	13
2.2.1. Método dos Mínimos Quadrados	14
2.2.2. Regressão Regularizada	15
2.3. Transferência de Aprendizagem	16
2.3.1. Transferência de aprendizagem X Aprendizagem multitarefas	16
2.3.2. Categorização da transferência de aprendizagem	17
2.3.3. Regressão Linear Regularizada Enviesada	18
2.4. Conclusão	18
3. IMPLEMENTAÇÃO	20
3.1. Ferramentas usadas para desenvolvimento	20
3.1.1. Biblioteca Scikit-learn	20
3.2. O que transferir	21
3.3. Como transferir	23
3.4. Conclusão	24
4. RESULTADOS	25
4.1. Experimentos	25
4.1.1. Métodos Avaliados	25
4.1.1.1. Modelo de origem sem transferência	25
4.1.1.2. Modelo de transferência com ajuste	25
4.1.1.3. Modelo de transferência sem ajuste	25
4.1.2. Metodologia de Avaliação	25
4.1.3. Comparação entre modelos	26
4.2. Conclusão	26
5. CONCLUSÃO FINAL	28
6. REFERÊNCIAS BIBLIOGRÁFICAS	29

# 1 – INTRODUÇÃO

## 1.1 Motivação

Algoritmos e otimizações em aprendizagem de máquina têm se mostrado interessantes quando se buscam soluções para alguns problemas da área de inteligência artificial. Entretanto, diante de várias soluções existentes, a busca de uma solução ótima para um determinado problema é fundamental quando um sistema é pensado para, além de resolver uma tarefa específica, ter um desempenho satisfatório onde a sua computação não seja considerada demasiadamente custosa.

Um pressuposto dos algoritmos de aprendizagem de máquina diz que os dados de treinamento devem possuir o mesmo espaço de características e a mesma distribuição dos dados de teste, mas em muitas situações do mundo real esta hipótese não se prende.

Portanto, em tais situações, o uso de transferência de aprendizagem se mostra uma solução de desempenho satisfatória, a qual pode ser usada tanto para problemas de regressão, classificação e agrupamento; quanto para problemas de mineração de dados.

## 1.2 Objetivo

O presente trabalho tem como objetivo a criação de um algoritmo para modelagem de mudança de contexto em problemas de regressão linear, feito através de técnicas de transferência de aprendizagem.

Como objetivos gerais, temos:

- Entender conceitos área de transferência de aprendizagem, partindo desde o grande tópico de aprendizagem de máquina, até a os subtipos de transferência comumente estudados e aplicados atualmente.
- Estudar as ferramentas computacionais apropriadas para o desenvolvimento de algoritmos na área de transferência de aprendizagem, no contexto desse trabalho.

Como objetivos específicos, temos:

- Propor e implementar um algoritmo que possa resolver um problema com dados reais, utilizando conceitos de transferência de aprendizagem e ferramentas computacionais ágeis para o desenvolvimento.

## 1.3 Estudo de caso

Este trabalho apresenta um estudo de caso comparativo sobre predições de performance em uma prova final de matemática para alunos que moram em zona rural. Três tipos de modelos de aprendizagem foram utilizados, no qual o primeiro fez uso de um conjunto de dados de treinamento deficiente, o segundo fez uso de transferência de aprendizagem, e o último fez uso um conjunto de dados de treinamento de um diferente contexto.

Experimentos com dados reais comprovaram que o melhor modelo a ser usado neste caso específico foi o modelo da transferência de aprendizagem entre contextos.

## **1.4 Organização**

No capítulo 2 será trazida uma fundamentação teórica base para o entendimento de todos os experimentos realizados por esse trabalho.

No capítulo 3 serão mostradas as ferramentas computacionais empregadas neste trabalho para a implementação do algoritmo proposto; assim como uma apresentação mais profunda de dois aspectos importantes da transferência de aprendizagem.

No capítulo 4 serão mostradas a estruturação dos experimentos realizados e os resultados obtidos.

No capítulo 5 serão mostradas as conclusões do trabalho, bem como eventuais melhorias e experimentos que poderão ser explorados em trabalhos futuros.

## 2 – FUNDAMENTAÇÃO TEÓRICA

Nos últimos anos, devido ao avanço tecnológico na área de inteligência artificial, cada vez mais algoritmos que lidam com aprendizagem de máquina são criados e aprimorados por diversos pesquisadores e entusiastas por todo o mundo. Alguns algoritmos, devido à sua complexidade e desempenho, são colocados em evidência em relação aos demais. Neste capítulo, serão abordados temas dentro do tópico de aprendizado de máquina como forma de fundamentação teórica para o entendimento dos experimentos que serão mostrados ao longo deste trabalho.

### 2.1 Aprendizagem de Máquina

“[...] Aprendizagem de máquina é um estudo sistemático de algoritmos de sistemas que melhoram o seu conhecimento ou performance com experiência” [1].

Em outras palavras, aprendizado de máquina pode ser caracterizado como toda e qualquer atividade na qual um computador pode ser capaz de interpretar uma situação e tomar uma decisão sem interferência direta de um ser humano. Um exemplo básico de decisão tomado por uma máquina é uma tarefa de classificação binária de reconhecimento de spams, na qual um modelo é programado de tal forma que o algoritmo, de forma autônoma, baseado em um conjunto prévio de dados, decide se um e-mail é ou não considerado spam [2].

Quando um algoritmo de aprendizagem é treinado usando um conjunto de dados, um modelo é criado para um dado problema. Tal modelo é feito por meio de uma inferência indutiva que busca prever eventos futuros e/ou derivar novos conhecimentos.

Em geral, aprendizagem de máquina é dividida em dois tipos: o supervisionado e o não supervisionado. A diferença entre eles está no formato do conjunto de dados. No primeiro tipo os dados possuem rótulos, e no segundo não. Isso faz com que o modelo gerado por uma aprendizagem supervisionada seja capaz de inferir uma função do conjunto de dados, o tornando apto para fazer previsões de eventos futuros [3].

Nos tópicos a seguir será feito um estudo sobre aprendizagem supervisionada e um estudo básico sobre aprendizagem não supervisionada, em vista que neste trabalho o foco será na transferência de aprendizagem supervisionada.

#### 2.1.1 Aprendizagem supervisionada

Aprendizagem supervisionada é baseada no mapeamento de determinados rótulos (saídas), dadas a entradas fornecidas para um determinado algoritmo. A figura 2.1.1 apresenta um esquema para esse tipo de aprendizagem.

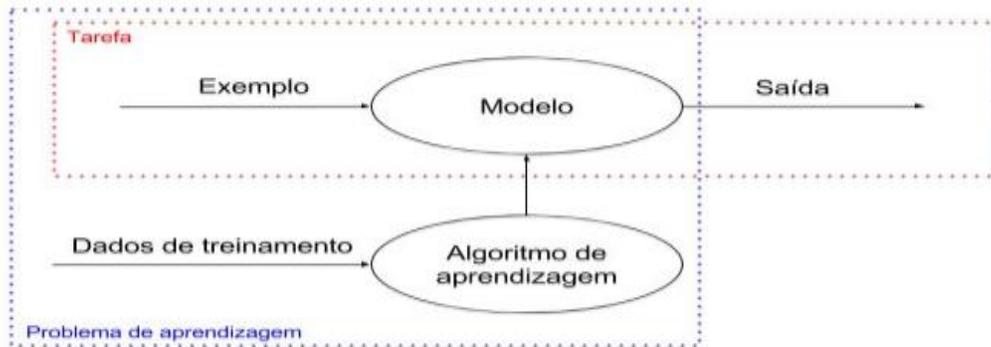


Figura 2.1.1.a - Esquema de uma aprendizagem supervisionada.

Os dados de treinamento utilizados neste tipo de aprendizagem são descritos através de um vetor de características e um rótulo associado a uma classe. O objetivo, ou tarefa, deste tipo de aprendizagem é construir um classificador (um modelo) capaz de determinar o rótulo/saída de novos exemplos. E este é construído a partir dos dados de treinamento e de um algoritmo de aprendizagem. Quando os rótulos são discretos, é dito como um problema de classificação, mas quando contínuos é dito regressão.

Depois da fase de treinamento do modelo, vem a fase de teste. Durante esta fase, dados também rotulados são utilizados para se fazer uma avaliação do modelo desenvolvido, calculando-se aqui uma taxa de erro (ou taxa de acerto).

O modelo de regressão linear é o exemplo mais simples desse tipo de aprendizagem, que usa o método dos mínimos quadráticos (que será abordado adiante neste trabalho) para minimização do erro quadrático da saída predita em relação ao valor real da saída [4].

Outro exemplo deste tipo de aprendizagem é a classificação linear [5], na qual o modelo busca dividir linearmente o espaço das características com um hiperplano de separação, como pode ser visto na figura 2.1.1.b.

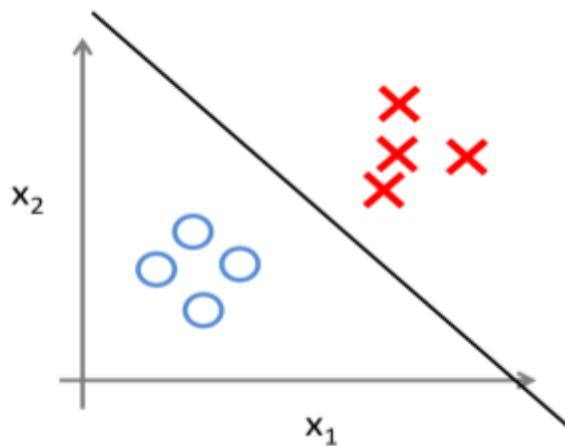


Figura 2.1.1.b - Exemplo de um modelo de classificação linear. Adaptado de: <http://oliviaklose.com/content/images/2015/02/2-supervised-vs-unsupervised-1.png>

### 2.1.2 Aprendizagem não supervisionada

Neste tipo de aprendizagem, os exemplos fornecidos não possuem rótulos, e a intenção é agrupá-los em grupos ou clusters, conforme figura 2.1.2. Após isso, uma análise de um especialista se faz necessária para entendimento do que estes agrupamentos vêm a significar no problema relacionado.

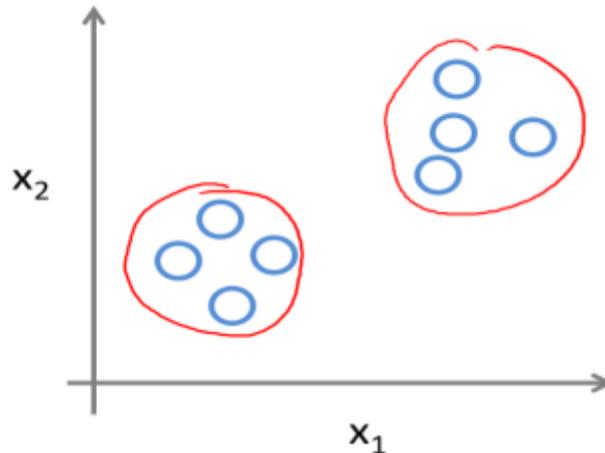


Figura 2.1.2 - Exemplo de clusters (agrupamentos) em dados não rotulados. Adaptado de: <http://oliviaklose.com/content/images/2015/02/2-supervised-vs-unsupervised-1.png>

Um método comumente conhecido neste tipo de aprendizagem é o algoritmo K-means, que visa agrupar um conjunto de observações não rotuladas em  $k$  grupos, considerando a média mais próxima entre eles e os centros de gravidades definidos para cada grupo. [6]

Não serão utilizados algoritmos de aprendizado não supervisionado neste trabalho, porém uma explanação foi feita logo acima para que se pudesse ser feita uma reflexão entre estes dois tipos de aprendizado existentes.

## 2.2 Regressão Linear

Regressão linear é um modelo estatístico que busca estimar o relacionamento entre uma variável dependente e uma ou mais variáveis independentes de forma linear. Um problema de aprendizagem regressiva é aquele que busca estimar uma função para a variável dependente através de observações e dados coletados. A figura 2.2 mostra um exemplo de regressão.

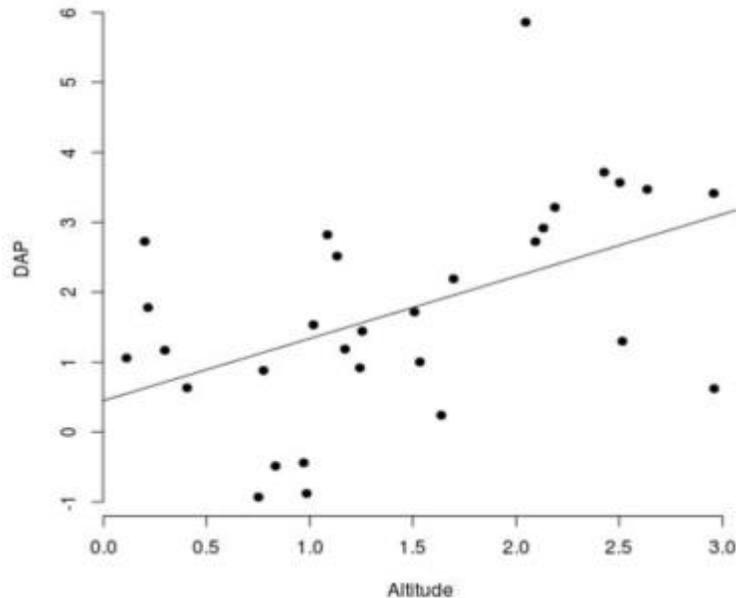


Figura 2.2 - Representação dos dados coletado e da função de estimação usando um algoritmo de regressão linear. Neste problema busca-se saber como a altitude (variável preditora) influencia no DAP (diâmetro a altura do peito) de uma espécie de árvore. Adaptado de: <http://recologia.com.br/tag/regressao-linear/>

### 2.2.1 Método dos Mínimos Quadrados

O método dos mínimos quadrados [7] busca encontrar uma solução global que minimiza a soma dos quadrados das diferenças entre o valor estimado e os dados observados. Ele é um modelo linear, ou seja, pode ser representado por uma função linear, que poderá ser aplicado tanto para problemas de regressão quanto para problemas de classificação.

Em problemas de regressão, o método busca estimar uma função linear  $f': X \rightarrow \mathfrak{R}$  considerando um conjunto de exemplos de treinamento  $(x_i, f(x_i))$  com  $n$  amostras. Os resíduos, ou erros, são dados através de  $\varepsilon_i = f(x_i) - f'(x_i)$ , e o método tem por objeto soma-los e tornar tal somatório  $S = \sum_{i=1}^n \varepsilon_i^2$  mínimo.

Assumindo que  $f'(x) = ax + b$ , podemos calcular os valores de  $a$  e  $b$  usando a noção de derivada onde  $\frac{\partial S}{\partial a} = 0$  e  $\frac{\partial S}{\partial b} = 0$ . Valores tais que formam o vetor de pesos  $\mathbf{w}$ , onde  $f'(x_i) = \mathbf{w}x_i$ . Com isso tem-se que:

$$a = \frac{n \sum_{i=1}^n x_i f(x_i) - \sum_{i=1}^n x_i \sum_{i=1}^n f(x_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (1)$$

$$b = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n f(x_i) - \sum_{i=1}^n x_i f(x_i) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (2)$$

Esta formulação da regressão considera apenas os resíduos na variável dependente e, por esse motivo, possui uma sensibilidade a ruídos que podem acontecer nas variáveis independentes. Os ruídos são pontos que estão longe da reta de regressão (função estimada) que podem causar *overfitting* - uma falta de generalização do modelo.

Apesar de possuir uma sensibilidade para ruídos, o método dos mínimos quadrados se mostra surpreendentemente eficiente para um modelo matematicamente tão simples. Um modo de olhar para isso é assumir que a função  $f$  é de fato linear, mas os valores de saída observados foram contaminados com algum ruído randômico.

### 2.2.2 Regressão regularizada

Quando a regressão é feita utilizando apenas o método dos mínimos quadrados, ela pode gerar uma solução instável altamente dependente dos dados de treinamento, que pode tender a uma situação de *overfitting*. Essa instabilidade pode ser resolvida por meio de métodos de regularização, que por sua vez adicionam algumas restrições ao vetor peso do modelo.

Ademais, a regularização reduz a complexidade do modelo, criando modelos mais simples que possuem maior generalização. Existem dois métodos de regularização que se destacam entre os demais em relação a aprendizagem de máquina voltada para problemas de regressão: o Ridge e o LASSO (*Least Absolute Shrinkage and Selection Operator*).

O método Ridge, também chamado de Regularização de Tikhonov, foi introduzido inicialmente pelo matemático Andrey Tikhonov em 1943 [8]. A regressão Ridge, regularização que usa o método Ridge, tem como característica a suavização de atributos correlacionados, os quais aumentam o ruído do modelo. Com a retirada de alguns atributos do modelo para a suavização, ele converge para uma solução mais estável. Portanto, atributos que contribuem menos para o poder preditivo do modelo são suavizados. Por exemplo, o atributo gênero em algum problema de regressão poderia ser suavizado caso não fosse muito relevante a escolha entre um homem ou mulher em determinada situação.

Busca-se então tornar o somatório  $S = \sum_{i=1}^n (f(x_i) - f'(x_i))^2$  mínimo, adicionando agora o elemento  $\lambda$ , chamado de parâmetro de suavização. Como resultado tem-se que:

$$\min_{\mathbf{w}} \sum_{i=1}^n (f(x_i) - \mathbf{w}x_i)^2 + \lambda \|\mathbf{w}\|^2 \quad (3)$$

Este tipo de regressão não tem o poder de zerar coeficientes, portanto estarão incluídos todos os coeficientes ou nenhum deles. Em contrapartida, a regularização LASSO tem essa capacidade, porque além do ajuste nos parâmetros, ela também seleciona variáveis automaticamente. Este segundo foi introduzido em 1996 por Robert Tibishirani em seu trabalho intitulado “*Regression Shrinkage and Selection via the lasso*” [9].

Com isso, o método Ridge não é recomendado para situações quando se têm um número exorbitante de atributos, pois representaria um desafio computacional. Já o LASSO é comumente escolhido para modelar situações onde o número de características é de milhões ou mais, pois se obtém uma solução esparsa onde atributos com coeficientes zero podem ser ignorados.

## 2.3 Transferência de Aprendizagem

Uma suposição na área de aprendizagem de máquinas é que muitos métodos funcionam bem se possuírem um bom conjunto de dados de treinamento para construção de um modelo, gerando assim bons resultados na etapa de teste. Todavia, em vários problemas do mundo real, nem sempre tal suposição se sustenta. Um exemplo é quando se têm o problema de classificação do grau do câncer em portadores de câncer de fígado, mas se obtém uma grande quantidade de dados de pacientes de câncer de pulmão. Ou seja, tem-se uma tarefa de classificação em um domínio de interesse, mas só se tem dados de treinamento suficientes em outro domínio de interesse, em que estes podem ser de um espaço de características diferentes ou seguirem uma distribuição de dados distinta.

Neste tipo de situação, a transferência de aprendizagem, se aplicada adequadamente, torna-se uma solução que melhora o desempenho de um modelo para um problema de aprendizagem. Ela pode ser aplicada tanto em problemas de aprendizagem supervisionada, classificação e regressão, quanto nos de não supervisionada e *clustering* [10].

### 2.3.1 Transferência de aprendizagem X Aprendizagem multitarefas

Diferentemente da aprendizagem de multitarefas (figura 2.3.1.a), em que se aprende todas as tarefas de origem e de destino simultaneamente, a transferência de aprendizagem (figura 2.3.1.b) tem um foco maior na tarefa de destino. Nela, as funções das tarefas de origem e destino são assimétricas, e elas não podem ser calculadas paralelamente.

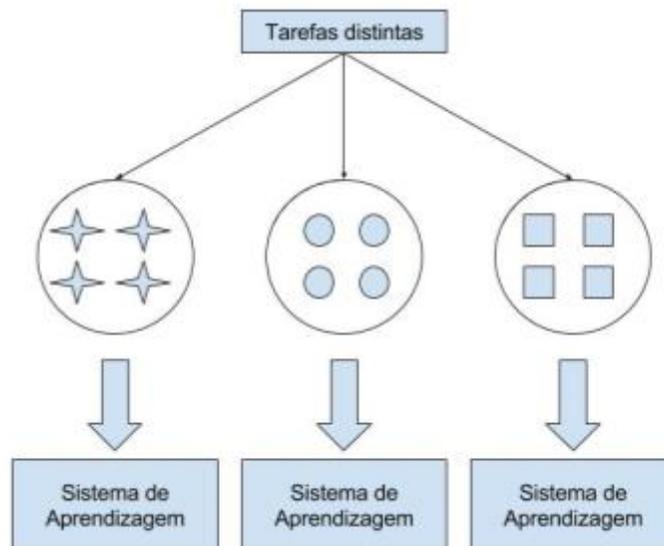


Figura 2.3.1.a - Esquemático de aprendizagem de multitarefas.

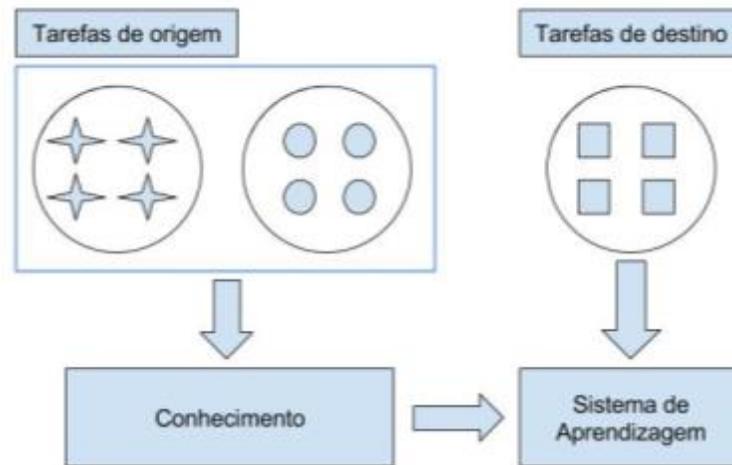


Figura 2.3.1.b - Esquemático da transferência de aprendizagem entre tarefas de origem e de destino.

Atualmente métodos de transferência de aprendizagem são aplicados em diversos campos, principalmente nas áreas de mineração de dados e aprendizagem de máquina [11][12][13].

### 2.3.2 Categorização da transferência de aprendizagem

Na transferência de aprendizagem, as seguintes três questões se mostram de grande relevância: O que transferir? Como transferir? Quando transferir?

“O que transferir” refere-se ao conhecimento aprendido das tarefas de origem. Deve-se ter em mente que alguns conhecimentos são específicos para alguns contextos, ou tarefas individuais e que outros podem ser comuns entre tarefas distintas. Estes, que são comuns, são aqueles que podem ajudar a melhorar o desempenho na tarefa de destino. Depois de saber o que pode ser transferido, a pergunta “Como transferir” é respondida através do desenvolvimento de algoritmos de aprendizagem para transferência de aprendizagem. Por fim, “Quando transferir” refere-se à quais situações a transferência pode ser realizada. Em situações onde o contexto de origem e de destino não possuem nenhuma relação entre si, considera-se inútil fazer a transferência, porque ela pode prejudicar o desempenho na tarefa de destino.

Levando em conta estas três questões apresentadas, a transferência de aprendizagem pode ser dividida em três subtipos: Transferência de aprendizagem indutiva, transdutora e não supervisionada [14].

A transferência de aprendizagem indutiva, a qual será utilizada neste trabalho, refere-se ao caso no qual os dados de treinamento das tarefas de origem e destino são rotulados. A transferência transdutora refere-se ao caso no qual os dados da tarefa de origem são rotulados e os da tarefa de destino não.

Por fim, a não supervisionada diz respeito ao caso onde nem os dados da tarefa de destino nem os da tarefa de origem são rotulados.

### 2.3.3 Regressão Linear Regularizada Enviesada

A regressão linear regularizada consiste em resolver o problema de otimização como visto na seção 2.2.2 em (3).

Porém, quando se trata de um problema de regressão no qual não há dados suficientes para criação de um modelo, o uso da expressão acima não se faz eficiente, e a solução vem com o uso da transferência de aprendizagem.

Mais ainda, para resolver tal situação, pode-se fazer uso da regressão linear regularizada enviesada [13][15], a qual também tem como objetivo minimizar o erro associado, mas agora o vetor de pesos  $\mathbf{w}$  é estimado de forma diferente.

Considerando um modelo  $f^O$  para tarefa de origem e um conjunto de exemplos de treinamento  $(x_i, f(x_i))$  com  $n$  amostras da tarefa de destino foram dados, busca-se modelar uma função  $f^D$  para a tarefa de destino. E, para isso, assume-se que  $f^D(x) = f^O(x) + g(x)$ , no qual  $g(x)$  é uma função de aproximação entre os dois contextos, o de destino e de origem.

Essa função de aproximação  $g$  é um modelo gerado através de uma regressão linear do erro associado  $\varepsilon_i = f(x_i) - f^O(x_i)$ . Ou seja,  $g(x_i) = \mathbf{w}'x_i$ , na qual  $\mathbf{w}'$  é o vetor de pesos associado ao modelo gerado. A partir daí tem-se que:

$$f^D(x_i) = f^O(x_i) + g(x_i)$$

$$f^D(x_i) = f^O(x_i) + \mathbf{w}'x_i$$

E em suma, fazendo-se as devidas substituições, tem-se que a regressão linear regularizada enviesada busca resolver o seguinte:

$$\min_{\mathbf{w}'} \sum_{i=1}^n (f(x_i) - f^O(x_i) - \mathbf{w}'x_i)^2 + \lambda \|\mathbf{w}'\|^2 \quad (4)$$

O trabalho realizado por Kuzborskij e Orabona em 2013 [13] mostrou que o algoritmo de transferência de aprendizagem baseado no método dos mínimos quadrados regularizados melhora sua generalização quando o modelo gerado da tarefa de origem funciona bem na tarefa de destino. Já no caso deste trabalho, a intenção é mostrar como se pode construir a função de aproximação  $g(x)$  tal que o algoritmo melhore em sua generalização, melhorando assim sua eficiência.

Um outro trabalho de Kuzborskij e Orabona [15] se concentraram na habilidade de generalização de uma grande família de tarefas de origem, para gerar um modelo para tarefa de destino usando um conjunto de hipóteses múltiplas fonte. O qual difere deste, onde busca-se gerar tal modelo usando apenas uma tarefa de origem.

## 2.4 Conclusão

Nesse capítulo foi feito um estudo sobre aprendizagem de máquina e se tornou possível compreender seu funcionamento, suas características e como ela se divide, tudo de acordo com algoritmos específicos. Se fez também notório o quão importante é ter uma boa quantidade de dados para a construção de modelos de aprendizagem, e que, quando isso não acontece, se faz

necessário o uso de técnicas de transferência de aprendizagem para se obter um resultado satisfatório na implementação e solução de um problema.

### 3 – IMPLEMENTAÇÃO

O foco deste trabalho consiste em otimizar a solução de um problema de regressão no qual não há dados suficientes para a geração de um modelo robusto. Tal problema pode ser resolvido com o uso de técnicas de transferência de aprendizagem, mais especificamente a técnica de transferência de modelos através de uma regressão linear regularizada enviesada.

Dado um domínio de origem onde se tem um bom modelo de predição já aprendido de forma confiável, o objetivo é reusar esse modelo em outros contextos (domínios) onde se tem poucos dados rotulados. Nesse caso, considerando as perguntas importantes de transferência de aprendizado temos: (1) o que transferir: nesse caso, um modelo aprendido em um domínio é usado como referência para aprendizado de novos modelos; (2) como transferir: nesse caso um novo modelo de predição para o domínio de destino é aprendido com os dados disponíveis, tendo o modelo do domínio de origem como referência. Para isso, a técnica de regressão enviesada aprenderá um modelo alvo similar ao modelo do domínio de origem, mas suficientemente ajustado para os dados disponíveis no domínio de destino. A suposição básica é que um bom modelo não poderia ser aprendido somente com os poucos dados etiquetados no domínio de destino.

Sobre a pergunta de quando transferir, não será investigado no trabalho, sendo proposto como trabalho futuro.

#### 3.1 Ferramentas usadas para desenvolvimento

Para a concretização da proposta desse trabalho de criação de um algoritmo para transferência de aprendizagem baseada em regressão linear regularizada enviesada, o conhecimento de como utilizar algumas tecnologias que envolvem programação numérica e científica, e análise sobre os atributos de um conjunto de dados, são fundamentais.

A seguir serão mostradas tecnologias que podem auxiliar no desenvolvimento da aplicação proposta por este trabalho.

##### 3.1.1 Biblioteca Scikit-learn

A linguagem Python é uma ferramenta muito poderosa para realização de cálculos matemáticos e experimentos científicos, mas para o contexto deste trabalho, que é voltado para aprendizagem de máquina, se faz necessário a inclusão de uma biblioteca que pudesse facilitar o desenvolvimento dos modelos de aprendizagem.

A SciKit-learn é uma biblioteca da linguagem Python para desenvolvimento na área de Aprendizagem de Máquina, e a mesma é um projeto *open source* de esforço comunitário, ou seja, qualquer pessoa é aberta a contribuir. Ela possui ferramentas simples e eficientes para mineração de dados e análise de dados e é comercialmente usada em aplicações do *Spotify* e da *Evernote* [16], por exemplo.

### 3.2 O que transferir

Para a realização de experimentos nesse trabalho, foi utilizado um conjunto de dados utilizado obtido no *UCI Machine Learning Repository* [17], repositório para testes na área de Aprendizagem de Máquina.

O conjunto de dados utilizado para realização destes experimentos foi um referente a performance de estudantes de escolas públicas, e este foi dividido em dois subconjuntos, o primeiro contendo alunos que moram em zona urbana e o outro contendo os alunos que moram em zona rural.

A tarefa de origem consiste em prever a nota de uma prova final de Matemática, que varia de 0 a 20, de um aluno que mora na zona urbana e possui um conjunto de treinamento com 307 instâncias. A tarefa de destino busca prever a mesma nota, sendo que, nesse caso, para um aluno que mora na zona rural, e está possui um conjunto de treinamento com 88 instâncias. O objetivo desse estudo de caso é transferir conhecimento de modelos de predição aprendizados no contexto de zona urbana para aprendizado de modelos de predição para o contexto de zona rural.

Os atributos para ambas as tarefas são os mesmos e consiste de 31 atributos diferentes, como pode ser visto na tabela abaixo.

<b>Nome</b>	<b>Tipo</b>	<b>Valores possíveis</b>
<b>Escola</b>	Binário	‘GP’ - Gabriel Pereira ou ‘MS’ - Mousinho da Silveira
<b>Sexo</b>	Binário	‘F’ - fêmea ou ‘M’ - masculino
<b>Idade</b>	Numérico	De 15 a 22
<b>Tamanho da família</b>		‘LE3’ - menor ou igual a 3 ou ‘GT3’ - maior que 3
<b>Estado de coabitação dos pais</b>	Binário	‘T’ - vivendo juntos ou ‘A’ - separado
<b>Educação materna</b>	Numérico	0 - nenhum, 1 - ensino primário (4ª série), 2 - 5º ao 9º ano, 3 - ensino secundário ou 4 - ensino superior
<b>Educação do pai</b>	Numérico	0 - nenhum, 1 - ensino primário (4ª série), 2 - 5º ao 9º ano, 3 - ensino secundário ou 4 - ensino superior
<b>Trabalho materno</b>	Nominal	‘Professor’, ‘saúde’ relacionado com cuidados, ‘servidor’ público (por exemplo, administrativo ou policial), ‘doméstico’ ou ‘outro’
<b>Trabalho do pai</b>	Nominal	‘Professor’, ‘saúde’ relacionado com cuidados, ‘servidor’ público (por exemplo, administrativo ou policial), ‘doméstico’ ou ‘outro’
<b>Razão para escolher esta escola</b>	Nominal	Perto de ‘casa’, ‘reputação’ da escola, ‘curso’ ou ‘outro’
<b>Guardião do estudante</b>	Nominal	‘Mãe’, ‘pai’ ou ‘outro’
<b>Tempo de viagem de casa para a escola</b>	Numérico	1 - < 15 min, 2 - 15 a 30 min, 3 - 30 min a 1 hora, ou 4 - > 1 hora
<b>Tempo de estudo semanal</b>	Numérico	1 - < 2 horas, 2 - 2 a 5 horas, 3 - 5 a 10 horas, ou 4 - > 10 horas

<b>Número de reprovações de classes passadas</b>	Numérico	n se $1 \leq n < 3$ , ou 4
<b>Apoio extra educacional</b>	Binário	Sim ou não
<b>Suporte educacional da família</b>	Binário	Sim ou não
<b>Aulas extra remuneradas de Matemática</b>	Binário	Sim ou não
<b>Atividades extracurriculares</b>	Binário	Sim ou não
<b>Escola maternal atendida</b>	Binário	Sim ou não
<b>Quer ter ensino superior</b>	Binário	Sim ou não
<b>Acesso à Internet em casa</b>	Binário	Sim ou não
<b>Possui um relacionamento romântico</b>	Binário	Sim ou não
<b>Qualidade das relações familiares</b>	Numérico	1 - muito ruim a 5 - excelente
<b>Tempo livre depois da escola</b>	Numérico	1 - muito ruim para 5 - muito bom
<b>Sair com amigos</b>	Numérico	1 - muito ruim para 5 - muito bom
<b>Consumo de álcool no dia útil</b>	Numérico	1 - muito ruim para 5 - muito bom
<b>Consumo de álcool no final de semana</b>	Numérico	1 - muito ruim para 5 - muito bom
<b>Estado de saúde atual</b>	Numérico	1 - muito ruim para 5 - muito bom
<b>Número de ausências escolares</b>	Numérico	De 0 a 93
<b>Nota da primeira prova de matemática</b>	Numérico	De 0 a 20
<b>Nota da segunda prova de matemática</b>	Numérico	De 0 a 20

Tabela 3.2.a - Atributos referentes aos modelos do contexto urbano e rural.

Para responder à pergunta de o que transferir, primeiramente foi realizada a discretização dos atributos, ou variáveis, que não eram numéricas para que se tornasse possível a construção dos modelos usando a biblioteca SciKit-learn. Isso foi realizado da seguinte forma: para as variáveis binárias, usou-se uma substituição dos valores pelos números 0 e 1, e para as variáveis nominais usou-se uma substituição dos nomes por valores de 1 até n, onde n era a quantidade de nomes possíveis para cada variável.

Em seguida, para garantir que no momento da transferência de aprendizagem não houvesse transferência de informação desnecessária, uma análise sobre os atributos de cada conjunto de dados foi realizada.

Primeiramente, modelos de regressão linear com regularização Ridge e com validação cruzada foram criados para cada uma das tarefas usando seus respectivos dados de treinamento. O parâmetro de suavização usado para regularização foi o valor padrão da função da biblioteca Scikit-learn.

Com os modelos prontos, pode-se obter então os valores dos coeficientes dos atributos para cada um dos contextos. E como pode ser observado na tabela 3.2.b, os pesos/coeficientes para os dois modelos de regressão possuem similaridades, o que nos permite fazer uso da transferência, mas também possuem diferenças, o que nos indica que para que haja uma transferência de conhecimento adequada se faz necessário um ajuste entre os contextos.

<b>Atributo</b>	<b>Coefficientes modelo rural</b>	<b>Coefficientes modelo urbano</b>
1	0.00263441	0.38063991
2	-0.28323377	-0.23759037
3	-0.06352453	-0.14497167
4	0.65185765	0.07568904
5	-0.43434555	-0.18440436
6	0.17610899	0.13658772
7	-0.01193663	-0.04122708
8	-0.08453278	0.04347891
9	0.05006465	-0.02985536
10	-0.16682638	0.14739606
11	0.38844266	-0.19318524
12	-0.0658485	-0.06744623
13	0.10658714	-0.03731173
14	-0.45336174	-0.33597736
15	-0.21540968	-0.15972896
16	-0.15226094	0.12763291
17	-0.15038664	-0.17400936
18	0.11696194	0.28127522
19	0.4218887	0.10891112
20	-0.77098005	-0.36409204
21	-0.0378228	0.00786156
22	0.34856178	0.36955614
23	-0.10363535	0.27316732
24	-0.03894499	0.0873867
25	0.05802819	-0.15123461
26	-0.03239654	-0.07879367
27	0.17529411	0.05444858
28	-0.10339493	0.03840839
29	0.02612918	0.03289096
30	0.29733627	0.40171653
31	0.51190295	0.49385645

Tabela 3.2.b - Coeficientes dos atributos referentes aos modelos do contexto urbano e rural.

### 3.3 Como transferir

Para responder à pergunta de como transferir, este trabalho fez uso da técnica de regressão linear regularizada enviesada. E, pela fórmula (6) em 2.3.3, ver-se que se faz necessário definir alguns parâmetros em ordem de responder esta questão.

Primeiramente, define-se a tarefa, ou contexto, de destino como a tarefa de prever a nota de uma prova final de matemática de um aluno que mora na zona rural. Como dito anteriormente, esta possui um conjunto de dados de treinamento de 88 instâncias.

Já a tarefa de origem, ou contexto de origem, busca prever a mesma nota, mas para um aluno que mora em zona urbana. E considerando seu conjunto de dados de treinamento de 308 instâncias, um modelo de regressão linear com regularização Ridge e com validação cruzada

$f^0$  foi treinado. O valor do parâmetro de suavização da regularização foi mais uma vez o valor padrão da função da biblioteca Scikit-learn.

Em seguida, define-se a função  $g$  como o modelo de erro entre os dois contextos, e para sua criação se faz necessário um conjunto de dados de treinamento. Este conjunto de dados será gerado da seguinte maneira.

1. Um subconjunto  $S$  de dados do contexto de destino de tamanho  $m$  é escolhido;
2. Todas as  $m$  instâncias  $(x_i, f(x_i))$  de  $S$  são treinadas no modelo  $f^0$  do contexto de origem e suas predições  $f^0(x_i)$  são coletadas;
3. O erro entre o valor real e o valor predito é calculado para todas as instâncias;  
$$\varepsilon_i = f(x_i) - f^0(x_i)$$
4. Gera-se o modelo de erro  $g$ , o qual é uma regressão linear com regularização Ridge e com validação cruzada, com conjunto de treinamento de tamanho  $m$  onde suas instâncias são  $(x_i, \varepsilon_i)$ .

Com o modelo de erro  $g$  treinado, se faz possível então resolver o problema de otimização de minimização do vetor de pesos  $\mathbf{w}'$ . Tornando possível o uso da transferência de aprendizagem para definir um modelo de aprendizagem para a tarefa de destino, como expresso na fórmula (5) da seção 2.3.3.

### 3.4 Conclusão

Esse capítulo buscou aplicar uma solução eficiente para um problema de regressão que não possui dados suficientes para criação de um modelo de aprendizagem robusto. Foi mostrado que soluções como o uso da linguagem Python, juntamente com a biblioteca SciKit-learn, são bem interessantes, uma vez que o desenvolvimento é simples e elas oferecem muitas funcionalidades na área de aprendizagem de máquina.

Também, neste capítulo, foi descrito todo o processo de escolha do o que transferir e como transferir, tornando possível a criação de um modelo de transferência de aprendizagem baseado em regressão regularizada.

Apesar do foco do uso das tecnologias descritas nesse capítulo estejam afeitas à solução do problema de predição de notas de alunos, essa descrição da implementação pode facilmente ser útil para o desenvolvimento de qualquer outro problema de regressão linear.

## 4 – RESULTADOS

Para a concretização da proposta deste trabalho, foi realizado uma série de experimentos, os quais fizeram uso dos dois conjuntos de dados citados na seção 3.2.

Buscou-se aqui fazer uma comparação entre o uso de um modelo de regressão linear regularizado para uma dada tarefa com um conjunto de treinamento deficiente, e o uso da transferência de aprendizagem entre duas tarefas distintas, onde a tarefa de destino possui seu conjunto de treinamento deficiente, e a tarefa de origem um conjunto robusto.

### 4.1 Experimentos

#### 4.1.1 Métodos avaliados

##### 4.1.1.1 Modelo de destino sem transferência

Um modelo de regressão linear regularizado com validação cruzada simples foi aprendido para a tarefa de destino com um subconjunto de seu conjunto de dados de treinamento. O valor do parâmetro de suavização usado na regularização Ridge foi o padrão definido pela biblioteca Scikit-Learn. O uso deste modelo é importante para se fazer uma comparação entre o uso de um modelo com dados deficientes e o uso da transferência de aprendizagem.

##### 4.1.1.2 Modelo de transferência com ajuste

Na seção 3.3 foram dadas as orientações de como se gerar um modelo de transferência de aprendizagem que usa regressão linear regularizado enviesado. Este modelo faz transferência com ajuste entre os dois contextos com o objetivo de mostrar a eficiência da transferência em relação ao uso de modelos deficientes, como citado em 4.1.1.1.

##### 4.1.1.3 Modelo de transferência sem ajuste

O modelo de transferência sem ajuste refere-se ao uso do modelo de regressão aprendido da tarefa de origem sem ajuste para a predição nos dados da tarefa de destino. A intenção em usar esse modelo é para fins de comparação entre o uso da transferência de aprendizagem com e sem ajuste entre diferentes contextos.

#### 4.1.2 Metodologia de Avaliação

O conjunto de dados original com 88 instâncias da tarefa de destino foi dividido igualmente em 10 *folds*, para que uma avaliação com validação cruzada fosse possível.

Desses 10 *folds*, 1 era selecionado como o conjunto de dados de teste e os outros 9 eram usados como o conjunto de dados de treinamento. Esse procedimento foi repetido 10 vezes de forma que cada um dos *folds* pudesse assumir o papel do conjunto de dados de teste uma vez.

A quantidade de dados de treinamento usada na aprendizagem dos dois modelos em 4.1.1.1 e 4.1.1.2 foi variada de forma a simular distintas situações onde diferentes números de

instâncias estão disponíveis na tarefa de destino. A quantidade variou de 1 até 79 instâncias para cada uma das iterações da validação cruzada.

### 4.1.3 Comparação entre os modelos

Para que houvesse uma comparação, cada conjunto de teste da tarefa de destino para cada iteração da validação cruzada foi treinado nos três modelos citados na seção 4.1.1, e o erro quadrático médio [18] entre a saída original e a saída predita foi calculado para cada instância de teste treinada.

Para os dois primeiros modelos, as médias dos erros foram calculadas para cada uma das 79 iterações correspondente as diferentes quantidades de dados de treinamento do contexto de destino disponíveis para cada iteração da validação cruzada. E, para o modelo de transferência sem ajuste, a média dos erros foi calculada considerando apenas os erros obtidos em cada uma das iterações da validação cruzada.

Como resultado, teve-se que em 100% dos casos com diferentes disponibilidades de dados de treinamento, o uso da transferência de aprendizagem com ajuste se mostrou mais eficiente do que o uso do um modelo de regressão sem transferência, pois proveu um resultado mais generalizado, com erros médios quadráticos menores. Também se observou que o uso do modelo de transferência sem ajuste foi melhor que o modelo com ajuste em 40% dos casos e melhor que o uso do modelo de regressão sem transferência em 100% dos casos. Essas conclusões podem ser observadas no gráfico mostrado na figura 4.1.3.

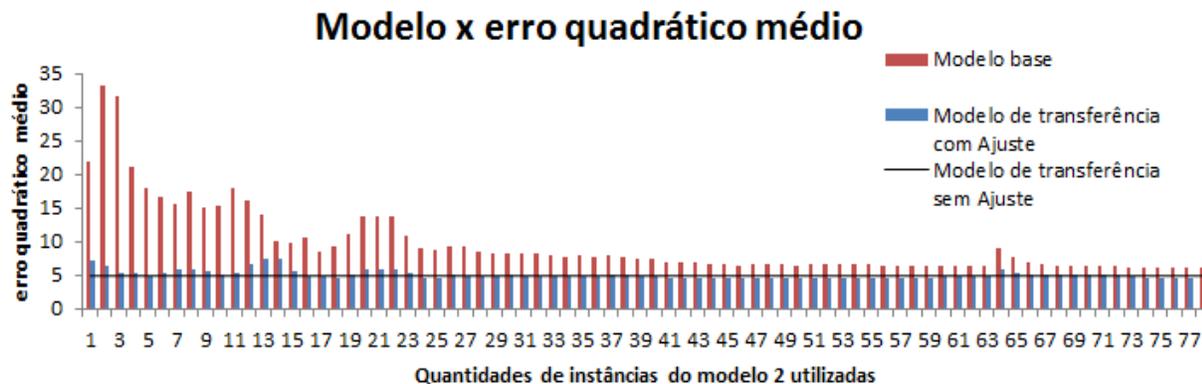


Figura 4.1.3 – Comparação entre o uso de Transferência de Aprendizagem entre duas tarefas, uso de uma Regressão Linear Regularizada com um conjunto de dados de treinamento deficiente (modelo base), uso do método da Regressão Linear Regularizada Enviesada (modelo de transferência com ajuste), e o uso do modelo de Regressão da tarefa de origem (modelo de transferência sem ajuste).

## 4.2 Conclusão

Neste capítulo foi descrito o processo de experimentação a fim de concretizar a proposta da implementação de um algoritmo de regressão linear regularizada enviesada.

Os experimentos no conjunto de dados obtidos no UCI Machine Learning Repository consolidaram o fato de que o algoritmo desenvolvido gerou um modelo coerente e generalizado para a tarefa de prever a nota de matemática de um aluno que mora em zona rural.

## 5 - CONCLUSÃO

Este trabalho buscou, por meio da implementação do algoritmo de regressão linear regularizada enviesada, prover uma solução para o problema de uma tarefa de regressão com um conjunto de dados de treinamento deficiente.

Experimentos comparativos foram realizados e eles comprovaram que para o cenário de predições de notas, o uso da transferência de aprendizagem com ajuste se mostrou mais eficiente pois obteve erros quadráticos médios menores entre os valores preditos e os valores originais dos dados de teste.

Como trabalho futuro, poderia ser feito um modelo de transferência de aprendizagem entre um conjunto de tarefas de origem e uma tarefa de destino, no qual se tem a ideia de um meta-aprendizado, e para tal, o modelo de erro para ajuste entre as tarefas deveria ser desenvolvido de tal forma que considera pesos diferentes para cada tarefa de origem.

Outro possível trabalho futuro seria expandir o algoritmo para transferência de aprendizagem entre tarefas classificação. No qual o modelo de ajuste entre as tarefas de origem e destino seria tal que modificaria ou não a escolha de uma classe predita.

## 6 - REFERÊNCIAS BIBLIOGRÁFICAS

- [1] FLACH, Peter. **Machine Learning: The Art and Science of Algorithms that Make Sense of Data**. New York: Cambridge, 2014.
- [2] AWAD, W.a. Machine Learning Methods for Spam E-Mail Classification. **International Journal Of Computer Science And Information Technology**, [s.l.], v. 3, n. 1, p.173-184, 28 fev. 2011. Academy and Industry Research Collaboration Center (AIRCC).
- [3] MOHRI, Mehryar; ROSTAMIZADEH, Afshin; TALWALKAR, Ameet. **Foundations of Machine Learning**. The MIT Press, 2012.
- [4] LEGENDRE, A. M.. **Nouvelles méthodes pour la détermination des orbites des comètes**. Firmin Didot, Paris, 1805.
- [5] YUAN, Guo-xun; HO, Chia-hua; LIN, Chih-jen. Recent Advances of Large-Scale Linear Classification. **Proceedings Of The Ieee**, [s.l.], v. 100, n. 9, p.2584-2603, set. 2012. Institute of Electrical and Electronics Engineers (IEEE).
- [6] KANUNGO, T. et al. An efficient k-means clustering algorithm: analysis and implementation. **Ieee Transactions On Pattern Analysis And Machine Intelligence**, [s.l.], v. 24, n. 7, p.881-892, jul. 2002. Institute of Electrical and Electronics Engineers (IEEE).
- [7] GAUSS, C. F.. **Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum**. 1809.
- [8] TIKHONOV, Andrey Nikolayevich. **On the stability of inverse problems**. Doklady Akademii Nauk: SSSR, 1943, p. 195–198.
- [9] TIBSHIRANI, Robert. 1996. Regression Shrinkage and Selection via the lasso. **Journal of the Royal Statistical Society**. v. 58, n. 1, p. 267–88, 1996. JSTOR.
- [10] PAN, Sinno Jialin; FELLOW, Qiang Yang. A Survey on Transfer Learning. **Ieee Transactions On Knowledge And Data Engineering**. p. 1345-1359. nov. 2010.
- [11] LUO, Ping; ZHUANG, Fuzhen; XIONG, Hui. Transfer Learning From Multiple Source Domains via Consensus Regularization. In: CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 17., 2008, Napa Valley. **Proceeding**. New York: Acm, 2008. p. 103 - 112.
- [12] HU, Shaofeng; REN, Jiangtao; ZHANG, Changshui. Domain Transfer via Multiple Sources Regularization. In: TSENG, Vincent; HO, Tu Bao; ZHOU, Zhi-hua. **Advances in Knowledge Discovery and Data Mining**. Tainan: Springer, 2014. p. 522-533.

- [13] ORABONA, Francesco; KUZBORSKIJ, Ilja. **Stability and Hypothesis Transfer Learning**. Disponível em: <<http://techtalks.tv/talks/stability-and-hypothesis-transfer-learning/58309/>>. Acesso em: 30 ago. 2016.
- [14] PAN, Sinno Jialin et al. A Survey on Transfer Learning. **Ieee Transactions On Knowledge And Data Engineering**, [s.l.], v. 22, n. 10, p.1345-1359, out. 2010. Institute of Electrical and Electronics Engineers (IEEE).
- [15] KUZBORSKIJ, Ilja; ORABONA, Francesco. Fast rates by transferring from auxiliary hypotheses. **Machine Learning**, [s.l.], v. 106, n. 2, p.171-195, 17 out. 2016. Springer Nature. <http://dx.doi.org/10.1007/s10994-016-5594-4>.
- [16] **SCIKIT-LEARN: Machine Learning in Python**. Disponível em: <<http://scikit-learn.org/stable/>>. Acesso em: 10 jan. 2017.
- [17] **MACHINE Learning Repository: Center for Machine Learning and Intelligent Systems**. Disponível em: <<http://archive.ics.uci.edu/ml/>>. Acesso em: 10 jan. 2017.
- [18] Venkatesh, Prasad K.. **Fundamentals of statistical signal processing: Estimation theory**: by Steven M. KAY. New Jersey: Englewood Cliffs, 1993.