



Bagging Predictors

Orivaldo Santana Jr.



Sumário

- Introdução
- Bootstrap
- Predição Numérica
- Ilustração de uma Regressão Linear
- Conclusão



Introdução

■ Origem do nome:

- Bagging = “**bootstrap aggregating**”

■ Bootstrap

- Gera vários conjuntos a partir de um conjunto de um único, por exemplo o de aprendizagem;
- Seleção dos elementos com reposição;
- Os novos conjuntos geralmente possuem o mesmo tamanho;
- Alguns elementos são repetidos;



Introdução

■ Agregação

- Em regressão pode ser feita com a média, mediana dos resultados do preditor;
- Em classificação pode ser feita por voto;

Introdução - Descrição formal

- $\mathcal{L}$: conjunto de aprendizagem
 - $\{(y_n, x_n), n = 1 \dots N\}$, com N exemplos;
 - y : pode ser labels ou respostas numéricas;
 - x_n : um vetor de regressores;
- $\varphi(x, \mathcal{L})$: preditor onde x é variável preditora;
- $\varphi_B(x) = av_B \varphi(x, \mathcal{L}^{(B)})$: Bagging para dados numéricos;
- $\mathcal{L}^{(B)}$: amostra bootstrap do conjunto de aprendizagem;



Bootstrap

- Existe duas abordagens básicas para estimar regressões bootstrapping;
- Na primeira é feito o ajuste do modelo baseada nos resíduos:
 - Ajusta o modelo de regressão com o intuito de estimar os betas;
 - “n” resíduos são obtidos, sendo “n” o tamanho da amostra;

Bootstrap

- Escolhe uma nova amostra de tamanho “ n ” destes resíduos com reposição e organiza em um vetor “ e^* ”;
- Com os regressores originais, o “ e^* ” e o “beta estimado” calcula o vetor de respostas bootstrap “ y^* ”;
- Novamente faz o ajuste do modelo desta vez com o “ y^* ”, produzindo a primeira estimativa bootstrap do vetor de coeficientes da regressão;
- Este procedimento pode ser repetido várias vezes para encontrar novos coeficientes;

Bootstrap

- A segunda abordagem (pares bootstrapping):
 - É usada:
 - Em situações onde existe dúvida sobre adequacidade da função de regressão;
 - A variância do erro não é constante;
 - Uma amostra bootstrap ($\mathcal{L}^{(B)}$) de tamanho “n” é gerada a partir do conjunto original ($\mathcal{L}$);
 - Ajusta o modelo utilizando esta amostra ($\mathcal{L}^{(B)}$) resultando a primeira estimativa bootstrap dos coeficientes de regressão;
 - Este procedimento pode ser repetido m vezes.



Bootstrap

- A segunda abordagem (pares bootstrapping):
 - Uma forma de selecionar m é observar a variabilidade do desvio padrão bootstrap com o aumento de m ;
 - Quando o desvio padrão bootstrap estabilizar, uma amostragem bootstrap de tamanho adequado é encontrada.

Embasamento estatístico

■ Predição Numérica:

- O bagging ou bootstrap funciona bem quando o preditor é instável, por exemplo redes neurais
- Se for estável: $\varphi_B \simeq \varphi(x, \mathcal{L})$, isto é, o preditor com bootstrap retornará resultados semelhantes ao preditor gerado diretamente do conjunto de aprendizagem;



Uma Ilustração de Regressão Linear

■ Seleção Forward de Variáveis

- Escolha progressiva das variáveis preditoras
 - $(x_1, \dots, x_m)$ escolhe x_k , onde $k = 1, \dots, m$ e m é o número de regressores
 - Escolha acumulativa
- Para cada escolha, calcular a soma dos quadrados dos resíduos (RSS);
- Melhor escolha é feita levando em conta o RSS;

Uma Ilustração de Regressão Linear

■ Estrutura da Simulação:

- $M = 30$, número de variáveis
- Tamanho da amostra é 60 exemplos
- x_m é gerado de uma distribuição normal com média zero.

$$y = \sum \beta_m x_m + \epsilon$$



Uma Ilustração de Regressão Linear

■ Estrutura da Simulação:

- Para cada conjunto de coeficientes (β) de regressão, o procedimento abaixo foi repetido 250 vezes
- Procedimento:
 - Calcular $\varphi_1(x), \dots, \varphi_M(x)$ com seleção forward de coeficientes e o erro médio quadrático de cada preditor.
 - Criar 50 replicas bootstrap da amostra e calcular $\varphi_1^{(B)}(x), \dots, \varphi_M^{(B)}(x)$ e a média dos erros de predição para eles.



Uma Ilustração de Regressão Linear

■ Estrutura da Simulação:

- Calcular a média dos erros do preditor com e sem bootstrap das 250 repetições;
- Os gráficos e os resultados serão explicados a seguir;



Uma Ilustração de Regressão Linear

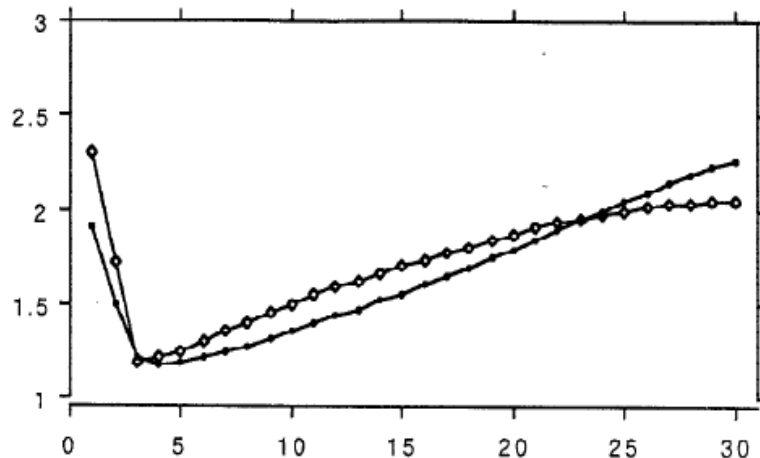
- Discussão dos Resultados da Simulação:
 - O mais exato preditor com “bagging” é no mínimo tão bom quanto o preditor sem “bagging”
 - Em todos os três gráficos existe um ponto que o preditor com “bagging” tem maior erro do que o sem “bagging”
 - Regressão linear usando todas as variáveis é um procedimento estável. A estabilidade diminui com o aumento do número de variáveis usadas no preditor.

Uma Ilustração de Regressão Linear

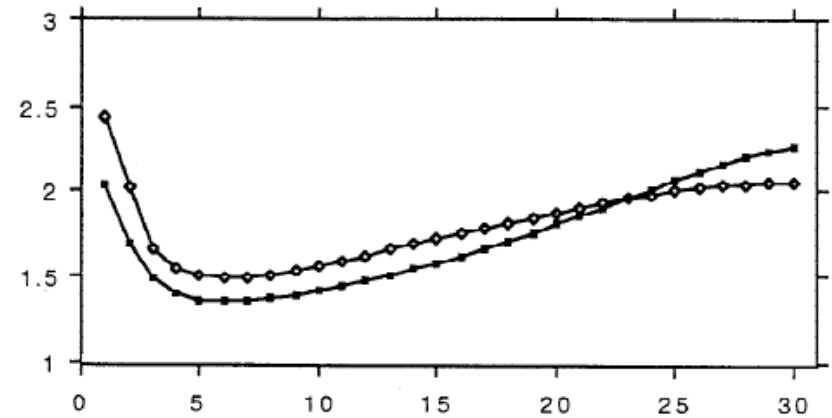
■ Discussão dos Resultados da Simulação:

◇ subset selection
— bagged subset selection

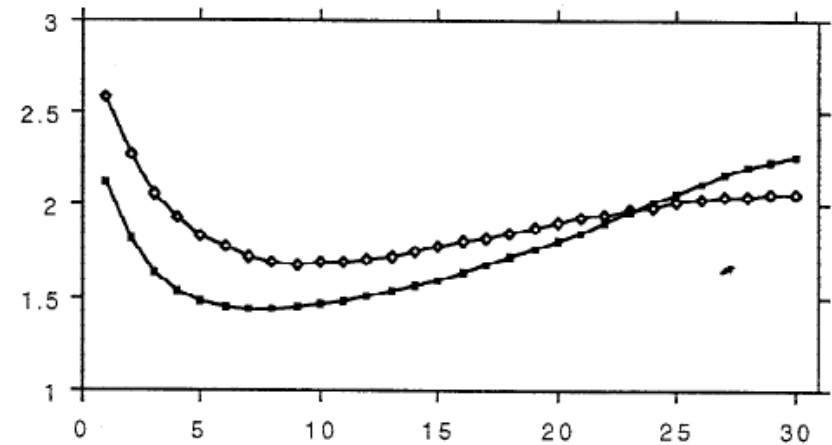
a) 3 nonzero coefficients



b) 15 nonzero coefficients



c) 27 nonzero coefficients





Conclusões

- Quantas replicas bootstrap são suficientes?
 - Nos experimentos de Breiman, 50 replicas bootstrap foram usadas para classificação e 25 para regressão;
 - Poucas replicas são necessários quando y é numérico;
 - Quanto maior o número de classes mais replicas são necessárias;



Conclusões

- Quantas replicas bootstrap são suficientes?
 - Em redes neurais a convergência é lenta, logo muitos replicas exigiram mais tempo de processamento;
 - Resultados da base de dados “waveform” com 10, 25, 50 e 100 replicas:
 - Taxa do preditor sem bagging foi 29,1
 - Com 10 replicas bootstrap a taxa foi de 21,8

Conclusões

- Quantas replicas bootstrap são suficientes?
 - A partir de 25 replicas a taxa não cresce significativamente:

Table 10. Bagged Misclassification Rates (%)

No. Bootstrap Replicates	Misclassification Rate
10	21.8
25	19.4
50	19.3
100	19.3



Conclusões

- Quanto grande deve ser o conjunto bootstrap?
 - Breiman usou o tamanho das replicas bootstrap igual ao tamanho do conjunto de aprendizagem.
 - Uma amostra bootstrap de tamanho N usa cerca de 63% dos exemplos.



Referências:

- Efron, B., & Tibshirani, R. (1993). “An Introduction to the Bootstrap”. Chapman and Hall.
- Breiman, L.(1996). Bagging predictors Machine Learning, Kluwer Academic Publishers, Volume 24, 123-140
- Montgomery, D. C.; Peck, E. A. & Vining, G. G. Wiley (ed.) Introduction to Linear Regression Analysis