1ª Lista de Exercícios - EST5534 Aprendizado de Máquina Estatístico

---

Exercício 1. Em um problema de regressão para um conjunto de dados $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n)\}$, considere que existe uma relação desconhecida entre a resposta e as preditoras dada por $y = r(\mathbf{x}) + \varepsilon$ e que funções de predição $g(\mathbf{x})$ foram usadas para predizer y em uma nova observação, em que se conhece apenas $\mathbf{x}$.

A perda quadrática é uma forma de medir o desempenho de modelos de regressão e é dada por $L(y, \widehat{g}(\mathbf{x})) = (y - \widehat{g}(\mathbf{x}))^2$, onde y é o valor alvo verdadeiro e $\widehat{g}(\mathbf{x})$ é a previsão do modelo.

Mostre que a decomposição do risco usando a perda quadrática em termos de viés e variância é dada por

$$\mathbb{E}[L(y, \widehat{g}(\mathbf{x}))] = \text{Viés}(\widehat{g}(\mathbf{x}))^2 + \text{Variância}(\widehat{g}(\mathbf{x})) + \sigma^2$$

Onde

- $\mathbb{E}[L(y, \widehat{g}(\mathbf{x}))]$ é o risco esperado, que representa o erro médio do modelo.
- $\text{Viés}(\widehat{g}(\mathbf{x})$ é o viés do modelo, definido como $\text{Viés}(\widehat{g}(\mathbf{x})) = \mathbb{E}[\widehat{g}(\mathbf{x})] - r(x)$.
- $\text{Variância}(\widehat{g}(\mathbf{x}))$ é a variância do modelo, definida como $\text{Variância}(\widehat{g}(\mathbf{x})) = \mathbb{E}[(\widehat{g}(\mathbf{x}) - E[\widehat{g}(\mathbf{x})])^2]$.
- σ^2 é a variância do ruído aleatório ε.

Você pode usar propriedades básicas de esperança, variância e a definição da perda quadrática para chegar a essa decomposição. Ao concluir a demonstração, explique como essa decomposição ilustra a relação entre viés, variância e erro irredutível no contexto de aprendizado de máquina.

Exercício 2. Em modelos supervisionados, considere um conjunto de dados de treinamento $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_n, y_n)\}$ é um vetor de características e y_i é a classe verdadeira de x_i. Considere que a previsão de um certo modelo foi $\hat{y}$ para um dado $\mathbf{x}$.

Considere as seguintes funções de perda:

- **Perda Quadrática:** $L(y, \hat{y}) = (y - \hat{y})^2$
- **Perda Absoluta:** $L(y, \hat{y}) = |y - \hat{y}|$
- **Perda 0-1:** $L(y, \hat{y}) = \begin{cases} 0 & \text{se } y = \hat{y} \\ 1 & \text{caso contrário.} \end{cases}$

Explique o contexto em que cada função de perda é mais adequada e justifique. Considere fatores como a sensibilidade a erros, o tipo de problema (regressão ou classificação), a presença de outliers nos dados e a interpretabilidade das previsões.

Exercício 3. Considerando as funções de perda do Exercício 2, proponha um breve estudo de simulação em um problema de regressão com uma variável preditora quantitativa contínua, em que se comparam métricas de ajuste como erro quadrático médio e erro absoluto médio, viés e variância estimados em diferentes modelos ajustados. Você deve considerar diversos modelos e produzir ao final gráficos de ajuste e uma tabela-resumo das quantidades obtidas. Repita para mais de uma variável preditora quantitativa contínua.

Nota: As variáveis preditoras devem ser fixadas em valores conhecidos, tal como uma função de regressão teórica escolhida e obter as respostas simuladas a partir de erros aleatórios gerados a partir de uma certa distribuição de probabilidade, por exemplo, normal padrão.

Exercício 4. Considere os dados `antracose` de Morettin e Singer (2022). As variáveis preditoras são `idade`, `tmunic`, `htrans`, `cargatabag`, `ses`, `densid` e `distmin` e a resposta `antracose`. A descrição das variáveis pode ser encontrada em Morettin e Singer (2022) e os códigos `02_ModelosdeRegularização.ipynb` podem auxiliar nas resoluções.

(1) Ajuste modelos linear sem e com regularização Ridge, Lasso e Elastic Net. Escolha os parâmetros de regularização via validação cruzada.
(2) Construa uma tabela com medidas de desempenho de todos os modelos ajustados e comente os resultados.

Exercício 5. (Morettin e Singer, 2022) Mostre que o estimador Ridge, $\widehat{\beta}_{\text{Ridge}}$, no modelo de regressão linear, é dado por

$$\widehat{\boldsymbol{\beta}}_{\text{Ridge}}(\lambda) = (I + \lambda(X^\top X)^{-1})^{-1}\widehat{\boldsymbol{\beta}}_{\text{MQ}}$$

em que $\widehat{\boldsymbol{\beta}}_{\text{MQ}}$ é o estimador de mínimos quadrados. Mostre que

$$\mathbb{E}(\widehat{\boldsymbol{\beta}}_{\text{Ridge}}(\lambda)) = (I + \lambda(X^\top X)^{-1})^{-1}\boldsymbol{\beta} \neq \boldsymbol{\beta},$$

ou seja, que $\widehat{\boldsymbol{\beta}}_{\text{Ridge}}(\lambda)$ é um estimador viesado.