Resposta ao Exercício sobre Decomposição do Erro Quadrático Esperado

(a) Explicação Conceitual de Bias^2, Variância e Ruído

A decomposição do erro quadrático esperado é uma ferramenta fundamental para entender os fatores que contribuem para o erro de um modelo preditivo. Considerando um problema de regressão onde desejamos prever uma variável alvo Y dado um vetor de entrada X, e temos um modelo $\hat{f}(X)$ estimado a partir de um conjunto de treinamento T, o erro quadrático esperado em um novo ponto x_0 pode ser decomposto da seguinte forma:

$$E[(Y - \hat{f}(x_0))^2 | X = x_0] = \sigma_\varepsilon^2 + [f(x_0) - E_T[\hat{f}(x_0)]]^2 + E_T[(\hat{f}(x_0) - E_T[\hat{f}(x_0)])^2]$$

onde:

- $\boldsymbol{\sigma_\varepsilon^2}$ representa o **ruído** (noise) ou erro irredutível. Este termo corresponde à variância inerente dos dados, ou seja, a variância da variável alvo Y em torno de sua verdadeira função $f(X)$. Mesmo um modelo perfeito não conseguiria reduzir esse erro, pois ele é uma propriedade dos dados. Assumindo $Y = f(X) + \varepsilon$ com $E[\varepsilon] = 0$ e $Var(\varepsilon) = \sigma_\varepsilon^2$, então $\sigma_\varepsilon^2 = Var(Y|X = x_0)$.
- $\mathbf{[f(x_0) - E_T[\hat{f}(x_0)]]^2}$ representa o **bias ao quadrado** ((bias)^2). O bias mede o quanto a média das predições do nosso modelo (treinado em diferentes conjuntos de treinamento T) se desvia da verdadeira função subjacente $f(x_0)$. Um modelo com alto bias faz suposições fortes sobre a forma da função, o que pode levar a um underfitting dos dados, ou seja, o modelo não captura as relações importantes entre as variáveis. $E_T[\hat{f}(x_0)]$ denota o valor esperado da predição $\hat{f}(x_0)$ sobre diferentes conjuntos de treinamento T.
- $\mathbf{E_T[(\hat{f}(x_0) - E_T[\hat{f}(x_0)])^2]}$ representa a **variância** (variance). A variância mede o quanto as predições do nosso modelo para um dado ponto x_0 variam entre diferentes conjuntos de treinamento T. Um modelo com alta variância é muito sensível às flutuações nos dados de treinamento, o que pode levar a um overfitting, ou seja, o modelo aprende o ruído presente nos dados de treinamento, generalizando mal para novos dados.

A **complexidade do modelo** tem um impacto significativo no bias e na variância. Tipicamente, existe um trade-off entre bias e variância:

- **Modelos Simples (Baixa Complexidade):** Tendem a ter um **alto bias** e **baixa variância**. Por exemplo:
  - **Regressão Polinomial de baixo grau:** Uma linha reta (grau $M = 1$) pode não conseguir capturar uma relação não linear (alto bias), mas diferentes conjuntos de treinamento provavelmente levarão a linhas retas semelhantes (baixa variância).
  - **k-NN com k grande:** Usar uma média de muitos vizinhos para fazer uma predição suaviza a resposta (alto bias) e torna a predição menos sensível a pontos de dados individuais no conjunto de treinamento (baixa variância).
  - **Regressão Ridge com λ grande:** Uma grande penalidade λ encolhe os coeficientes para zero, simplificando o modelo (alto bias) e tornando-o menos sensível aos dados de treinamento (baixa variância).
- **Modelos Complexos (Alta Complexidade):** Tendem a ter um **baixo bias** e **alta variância**. Por exemplo:

– **Regressão Polinomial de alto grau:** Um polinômio de grau alto ($M = 9$) pode se ajustar muito bem aos dados de treinamento, inclusive ao ruído (baixo bias no treinamento), mas pequenas mudanças no conjunto de treinamento podem levar a curvas muito diferentes (alta variância), resultando em um mau desempenho em novos dados (overfitting).

– **k-NN com k pequeno:** Usar apenas um vizinho para fazer uma predição torna o modelo muito flexível (baixo bias no treinamento) mas também muito sensível a pontos de dados individuais no conjunto de treinamento (alta variância).

– **Regressão Ridge com λ pequeno:** Uma pequena penalidade λ permite que o modelo se ajuste mais livremente aos dados de treinamento (baixo bias), mas também o torna mais suscetível ao ruído nos dados de treinamento (alta variância).

O objetivo é encontrar um modelo com uma complexidade ótima que minimize o erro quadrático esperado total, alcançando um bom equilíbrio entre bias e variância.

(b) Componente Prático: Escolha de Dataset e Ajuste de Modelos

Para este componente prático, você deve:

1. Escolher um dataset de regressão simples. Você pode:

   - Gerar dados sintéticos seguindo exemplos da disciplina, como uma relação não linear com adição de ruído Gaussiano. Por exemplo, gerar pontos (x_i, t_i) onde $t_i = \sin(2\pi x_i) + \varepsilon_i$, com $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ e x_i uniformemente distribuídos no intervalo de 0 a 1.
   - Usar um dataset padrão pequeno de uma biblioteca como scikit-learn (para Python) ou outras.

2. Ajustar modelos de complexidade variável ao seu dataset escolhido. Considere as seguintes opções (escolha uma ou mais):

   - **Regressão Polinomial:** Ajuste modelos com graus polinomiais crescentes M (e.g., $M = 1, 2, 3, 5, 10$).
   - **k-Nearest Neighbors (k-NN):** Ajuste modelos com valores decrescentes de k (e.g., $k = 1, 3, 5, 10, \text{tamanho do dataset}$).
   - **Regressão Ridge:** Ajuste modelos com valores decrescentes de λ (o coeficiente de regularização). Note que λ controla a força da regularização.

(c) Componente Prático: Estimação Empírica e Plotagem de Bias^2 e Variância

Para este componente prático, você deve:

1. **Estimar empiricamente o bias ao quadrado e a variância** para cada nível de complexidade do modelo ajustado na parte (b). Uma maneira de fazer isso é:

   - **Divisão Treino/Teste Repetida:** Divida seu dataset repetidamente em conjuntos de treinamento e teste (e.g., 80
   - **Estimação do Bias:** Para cada ponto de avaliação, calcule a média de todas as predições dos modelos treinados. O bias é então estimado como o quadrado da diferença entre essa média e o valor verdadeiro (se conhecido para dados sintéticos) ou uma "boa" estimativa da verdade subjacente.
   - **Estimação da Variância:** Para cada ponto de avaliação, calcule a variância das predições dos diferentes modelos treinados. A média dessas variâncias sobre os pontos de avaliação fornece uma estimativa da variância do modelo para um dado nível de complexidade.
   - **Bootstrap:** Se você tem um dataset fixo, você pode gerar múltiplos conjuntos de treinamento bootstrap (amostragem com reposição). Treine o modelo em cada conjunto bootstrap e siga um procedimento similar ao da divisão treino/teste repetida para estimar bias e variância.

2. **Plote** como o bias ao quadrado e a variância mudam com a complexidade do modelo. O eixo x do seu gráfico deve representar a complexidade do modelo (e.g., grau M, $1/k$, ou $1/\lambda$), e o eixo y deve representar os valores estimados do bias ao quadrado e da variância. Você deve ter duas curvas no seu gráfico.

3. **Discuta** se os seus resultados empíricos estão alinhados com o trade-off teórico entre bias e variância discutido na parte (a). Você deve observar que, à medida que a complexidade do modelo aumenta, o bias tende a diminuir e a variância tende a aumentar, e vice-versa. Identifique a região de complexidade onde você pode obter um bom equilíbrio.

Este componente prático requer implementação computacional para gerar os dados, ajustar os modelos, realizar as simulações e gerar os gráficos. As referências nos seus materiais didáticos (como as Figuras 1.6, 4.7, 7.1 e as discussões sobre overfitting e underfitting nos Capítulos 1, 3, 4 e 7) devem fornecer exemplos e intuições úteis para realizar esta parte do exercício.