INSTITUTO DE INFORMÁTICA
Universidade Federal de Goiás

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
TÓPICOS ESPECIAIS EM FUNDAMENTOS DE COMPUTAÇÃO – MATEMÁTICA E ESTATÍSTICA PARA CIÊNCIA DE DADOS
PROF. DR. ROMMEL MELGAÇO BARBOSA

SEMINÁRIOS

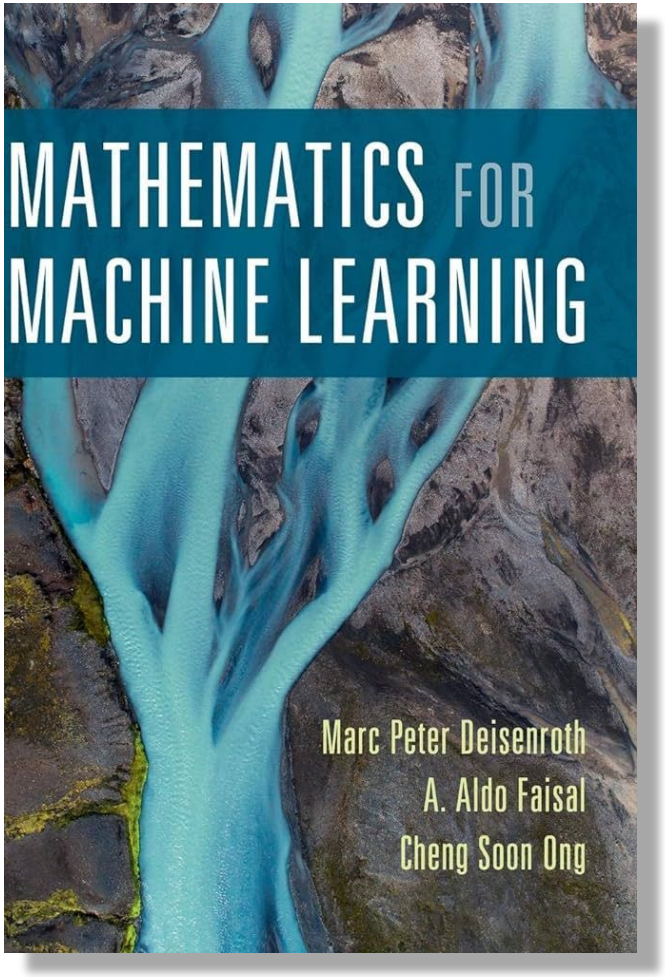
REGRESSÃO LINEAR

João Gabriel Junqueira da Silva
Manoel Verissimo dos Santos Neto
Rafael Rodrigues Silva

Junho/2024

INSTITUTO DE INFORMÁTICA UFG

UFG UNIVERSIDADE FEDERAL DE GOIÁS



Capítulo 09

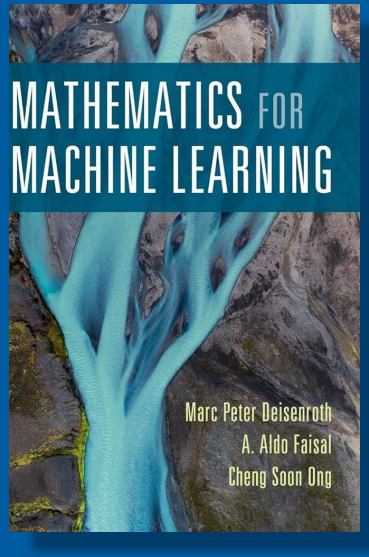
Regressão Linear

- 9.1 Formulação do Problema
- 9.2 Estimativa de Parâmetros
  - 9.2.1 Estimativa de Máxima Verossimilhança
  - 9.2.2 Overfitting em Regressão Linear
  - 9.2.3 Estimativa Máxima a Posteriori
  - 9.2.4 Estimativa MAP como Regularização
- 9.3 Regressão Linear Bayesiana
  - 9.3.4 Distribuição Posterior
  - 9.3.5 Previsões Posteriores
  - 9.3.5 Calculando a Probabilidade Marginal
- 9.4 Máxima Probabilidade como Projeção Ortogonal

Capítulo 09

Regressão Linear

9.1 Formulação do Problema



9.1 Formulação do Problema

Definição: A formulação do problema de regressão linear envolve definir a relação entre uma variável dependente ***y*** e uma ou mais variáveis independentes ***X***. O objetivo é encontrar os coeficientes de um modelo linear que melhor se ajuste aos dados observados.

9.1 Formulação do Problema

Modelo de Regressão Linear Simples

Definição: Na regressão linear simples, modelamos a relação entre uma única variável independente x e a variável dependente y usando a equação:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Onde:

- ***y*** é a variável dependente (resposta).
- ***x*** é a variável independente (preditor).
- β_0 é o intercepto da reta.
- β_1 é o coeficiente angular (slope).
- ϵ é o termo de erro aleatório, que segue uma distribuição normal $\epsilon \sim N(0, \sigma^2)$.

9.1 Formulação do Problema

Modelo de Regressão Linear Múltipla

Definição: Na regressão linear múltipla, modelamos a relação entre várias variáveis independentes $x_1, x_2, \ldots, x_p$ e a variável dependente y usando a equação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

Ou, em notação matricial:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Onde:

- y é o vetor das observações da variável dependente (n x 1).
- X é a matriz das observações das variáveis independentes (n x (p+1)), com a primeira coluna sendo composta por 1s (para o intercepto).
- β é o vetor dos coeficientes do modelo ((p+1) x 1).
- ϵ é o vetor dos termos de erro (n x 1).

9.1 Formulação do Problema

Suposições do Modelo de Regressão Linear

Para que o modelo de regressão linear produza estimativas válidas e úteis, algumas suposições devem ser satisfeitas:

1. **Linearidade:** A relação entre a variável dependente e as variáveis independentes é linear.
2. **Independência:** As observações são independentes entre si.
3. **Homocedasticidade:** A variância dos erros é constante para todos os valores das variáveis independentes.
4. **Normalidade dos Erros:** Os erros do modelo são normalmente distribuídos com média zero.

9.1 Formulação do Problema

Função de Custo

A função de custo na regressão linear é a soma dos quadrados dos resíduos (SSE):

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Onde:

- y_i são os valores observados.
- $\hat{y}_i$ são os valores preditos pelo modelo.

9.1 Formulação do Problema

Solução pelo Método dos Mínimos Quadrados

Para encontrar os coeficientes que minimizam a soma dos quadrados dos resíduos, usamos o método dos mínimos quadrados. Em notação matricial, a solução é obtida resolvendo a equação normal:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Essa equação nos fornece os estimadores dos coeficientes que melhor se ajustam aos dados, minimizando a soma dos quadrados dos resíduos.

9.1 Formulação do Problema

Interpretação dos Coeficientes

- $\boldsymbol{\beta_0}$ (intercepto): Representa o valor esperado de y quando todas as variáveis independentes são iguais a zero.
- $\boldsymbol{\beta_j}$ (coeficiente de x_j): Representa a mudança esperada em $\boldsymbol{y}$ para uma unidade de mudança em $\boldsymbol{x_j}$, mantendo todas as outras variáveis constantes.

9.1 Formulação do Problema

Avaliação do Modelo

Para avaliar a qualidade do ajuste do modelo, utilizamos métricas como o R^2 e o erro padrão dos resíduos (SE):

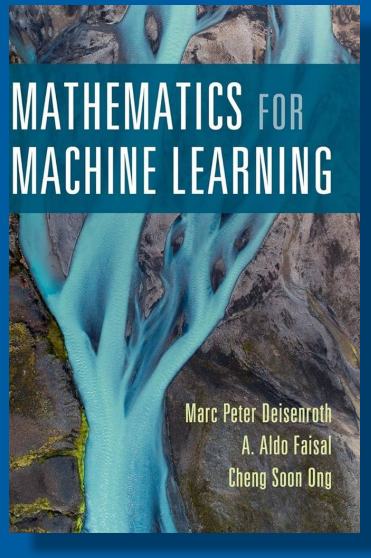
- **R^2 (coeficiente de determinação):** Mede a proporção da variância total em y explicada pelo modelo. É calculado como:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- **Erro padrão dos resíduos (SE):** Fornece uma medida da precisão das previsões do modelo. É calculado como:

$$SE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p-1}}$$

9.2 Estimativa de Parâmetros



9.2 Estimativa de Parâmetros

A estimativa de parâmetros na regressão linear envolve encontrar os coeficientes $\boldsymbol{\beta}$ que melhor descrevem a relação entre as variáveis independentes $\boldsymbol{X}$ e a variável dependente $\boldsymbol{y}$. O método mais comum para essa estimativa é o método dos mínimos quadrados ordinários (OLS - Ordinary Least Squares).

9.2 Estimativa de Parâmetros

Modelo de Regressão Linear Simples

Para a regressão linear simples, o modelo é:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Queremos encontrar os valores de β_0 e β_1 que minimizam a soma dos quadrados dos resíduos (SSE):

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Onde:

- y_i são os valores observados da variável dependente.
- $\hat{y}_i$ são os valores preditos pelo modelo:

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

9.2 Estimativa de Parâmetros

Estimadores dos Coeficientes

Os estimadores dos coeficientes $\boldsymbol{\beta_0}$ e $\boldsymbol{\beta_1}$ são obtidos resolvendo-se o sistema de equações normais derivadas da minimização da SSE.

Para β_1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Para β_0:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Onde $\bar{x}$ e $\bar{y}$ são as médias das observações de **x** e **y**, respectivamente.

9.2 Estimativa de Parâmetros

Derivação dos Estimadores

Vamos derivar as fórmulas para os estimadores usando o método dos mínimos quadrados. Primeiro, reescrevemos a SSE:

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Para minimizar a SSE, derivamos em relação a $\boldsymbol{\beta_0}$ e $\boldsymbol{\beta_1}$ e igualamos a zero:

$$\frac{\partial \text{SSE}}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial \text{SSE}}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

Resolvendo essas duas equações simultaneamente, obtemos as fórmulas para $\boldsymbol{\hat{\beta}_0}$ e $\boldsymbol{\hat{\beta}_1}$.

9.2 Estimativa de Parâmetros

Modelo de Regressão Linear Múltipla

Para a regressão linear múltipla, o modelo é:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

y = **X**β + ϵ

Onde:

- **y** é o vetor das observações da variável dependente (n x 1).
- **X** é a matriz das observações das variáveis independentes (n x (p+1)), com a primeira coluna sendo composta por 1s (para o intercepto).
- $\boldsymbol{\beta}$ é o vetor dos coeficientes do modelo ((p+1) x 1).
- $\boldsymbol{\epsilon}$ é o vetor dos termos de erro (n x 1).

9.2 Estimativa de Parâmetros

Estimadores dos Coeficientes

Os coeficientes são estimados minimizando a soma dos quadrados dos resíduos em notação matricial:

$$\text{SSE} = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

Para encontrar os estimadores, derivamos a SSE em relação a β e igualamos a zero:

$$\frac{\partial \text{SSE}}{\partial \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

Resolvendo para ***β***, obtemos:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

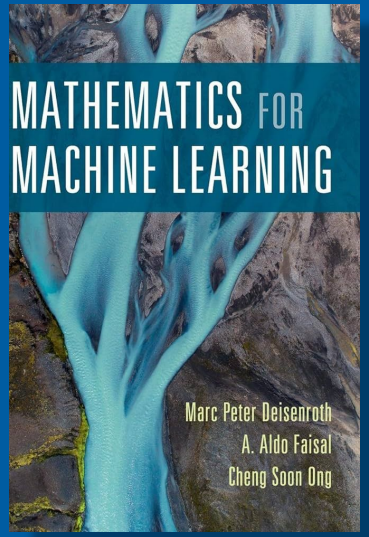
9.2 Estimativa de Parâmetros

Propriedades dos Estimadores

Os estimadores obtidos pelo método dos mínimos quadrados possuem algumas propriedades importantes:

- **Não viesados:** Os estimadores são não viesados, ou seja, $E(\hat{\beta}) = \beta$.
- **Variância mínima:** Entre todos os estimadores lineares não viesados, os estimadores de mínimos quadrados têm a menor variância.
- **Distribuição:** Se os erros ϵ são normalmente distribuídos, então os estimadores $\hat{\beta}$ também são normalmente distribuídos.

9.2.1 Estimativa de Máxima Verossimilhança



9.2.1 Estimativa de Máxima Verossimilhança

Para encontrar os parâmetros que maximizam a verossimilhança, utilizamos a técnica de máxima verossimilhança:

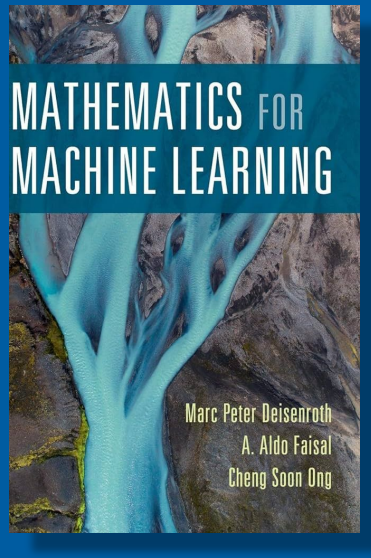
$$\boldsymbol{\theta}_{\mathrm{ML}} \in \arg\max_{\boldsymbol{\theta}} p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) \,.$$

A solução analítica para a estimativa de máxima verossimilhança é:

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \,.$$

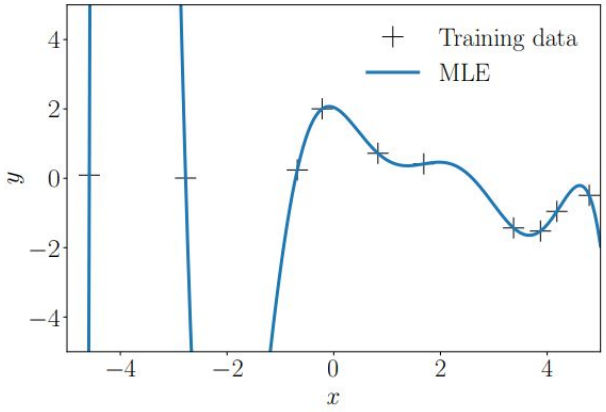
onde **X** é a matriz de design que contém os vetores de entrada $\boldsymbol{x}_n$

9.2.2 Overfitting em Regressão Linear

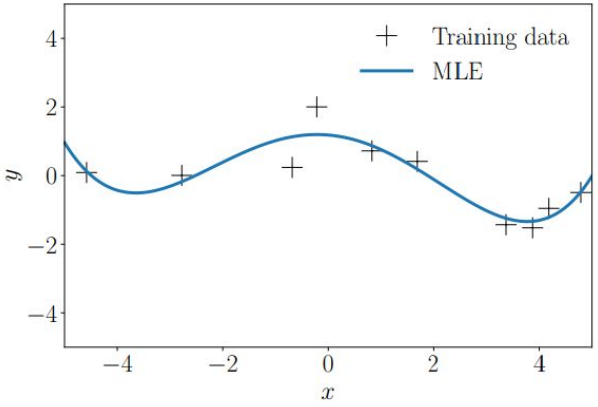


9.2.2 Overfitting em Regressão Linear

Overfitting ocorre quando o modelo se ajusta muito bem aos dados de treinamento, mas não generaliza bem para novos dados. Isso geralmente acontece quando o modelo é muito complexo.



(a) Overfitting



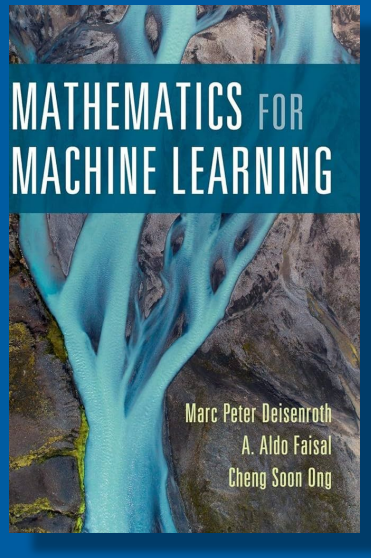
(c) Fitting well.

9.2.2 Overfitting em Regressão Linear

Técnicas para Evitar Overfitting

- **Regularização:** Adicionar um termo de penalização na função de custo, como na estimativa MAP.
- **Validação Cruzada:** Utilizar técnicas de validação cruzada para avaliar o desempenho do modelo em diferentes subconjuntos dos dados.
- **Seleção de Modelo:** Escolher o modelo mais simples que ainda explica bem os dados, utilizando critérios como o erro quadrático médio (MSE) ou o erro quadrático médio da raiz (RMSE).

9.2.3 Estimativa Máxima a Posteriori



9.2.3 Estimativa Máxima a Posteriori

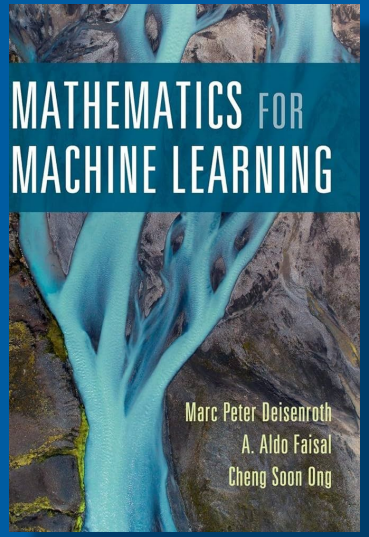
A estimativa de máxima verossimilhança é propensa a overfitting. Para mitigar esse efeito, podemos impor uma distribuição priori p(θ) sobre os parâmetros. A distribuição priori codifica os valores plausíveis dos parâmetros antes de observar os dados. A fórmula da posteriori é dada por:

$$p(\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}) = \frac{p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathcal{Y} \mid \mathcal{X})} .$$

O vetor de parâmetros θ_MAP que maximiza a posteriori é a estimativa MAP. Para encontrar a estimativa MAP, minimizamos a função:

$$\boldsymbol{\theta}_{\mathrm{MAP}} \in \arg\min_{\boldsymbol{\theta}} \{ -\log p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) - \log p(\boldsymbol{\theta}) \} .$$

9.2.4 Estimativa MAP como Regularização



9.2.4 Estimativa MAP como regularização

Além de impor uma distribuição priori sobre os parâmetros, é possível mitigar o overfitting penalizando a magnitude dos parâmetros por meio da regularização. Na regularização de mínimos quadrados, consideramos a função de perda:

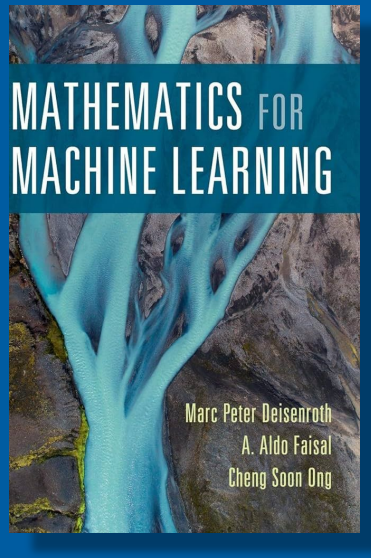
$$\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\theta}\|^2 + \lambda \|\boldsymbol{\theta}\|_2^2$$

O termo de regularização λ controla a rigidez da regularização. A minimização desta função de perda fornece a solução:

$$\boldsymbol{\theta}_{\text{RLS}} = (\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda \boldsymbol{I})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{y}$$

Essa solução é idêntica à estimativa MAP para $\lambda = \sigma^2 / b^2$, onde σ^2 é a variância do ruído e b^2 é a variância da priori Gaussiana isotrópica.

9.3 Regressão Linear Bayesiana



9.3 Regressão Linear Bayesiana

Na regressão linear Bayesiana, não buscamos uma estimativa pontual dos parâmetros, mas sim a distribuição posterior completa dos parâmetros. A posteriori dos parâmetros é dada por:

$$p(\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}) = \mathcal{N}\left(\boldsymbol{\theta} \mid \boldsymbol{m}_N, \boldsymbol{S}_N\right)$$

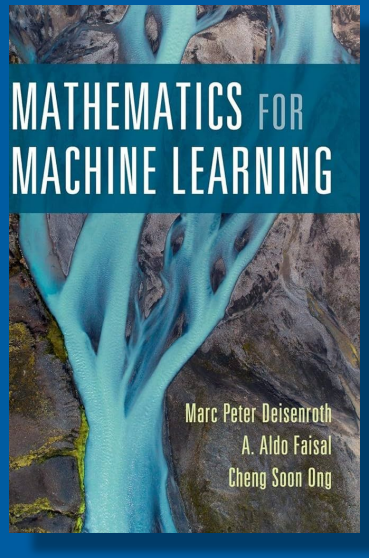
Onde Sn e Mn são calculados como:

$$\boldsymbol{S}_N = (\boldsymbol{S}_0^{-1} + \sigma^{-2}\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1},$$
$$\boldsymbol{m}_N = \boldsymbol{S}_N(\boldsymbol{S}_0^{-1}\boldsymbol{m}_0 + \sigma^{-2}\boldsymbol{\Phi}^\top\boldsymbol{y})$$

Para fazer previsões, usamos a distribuição posterior e integramos sobre todos os parâmetros plausíveis:

$$p(y^* \mid X, Y, x^*) = N(y^* \mid \phi(x^*)^T m_N, \phi(x^*)^T S_N \phi(x^*) + \sigma^2)$$

9.3.1 Modelo


MATHEMATICS FOR
MACHINE LEARNING
Marc Peter Deisenroth
A. Aldo Faisal
Cheng Soon Ong

9.3.1 Modelo

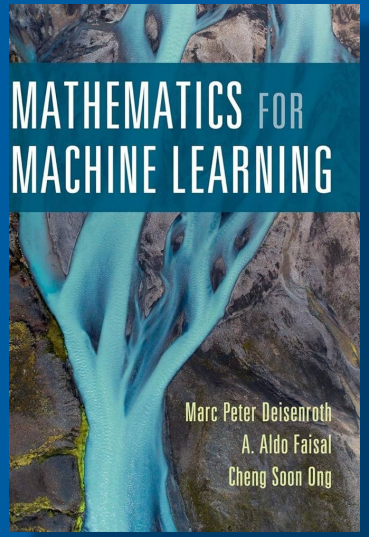
Na regressão linear Bayesiana, consideramos o modelo:

$$p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{S}_0)\,,$$
$$p(y \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(y \mid \boldsymbol{\phi}^\top(\boldsymbol{x})\boldsymbol{\theta}, \sigma^2)$$

Isso transforma o vetor de parâmetros em uma variável aleatória e permite escrever o modelo probabilístico completo:

$$p(y, \boldsymbol{\theta} \mid \boldsymbol{x}) = p(y \mid \boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})$$

9.3.2 Previsões anteriores



9.3.2 Previsões anteriores

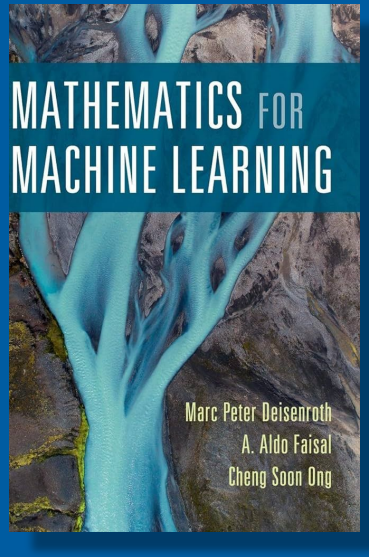
Para fazer previsões em x∗, integramos θ e obtemos:

$$p(y_* \mid \boldsymbol{x}_*) = \int p(y_* \mid \boldsymbol{x}_*, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}}[p(y_* \mid \boldsymbol{x}_*, \boldsymbol{\theta})]$$

Com uma priori Gaussiana em θ, a distribuição preditiva é Gaussiana:

$$p(y_* \mid \boldsymbol{x}_*) = \mathcal{N}\left(\boldsymbol{\phi}^\top(\boldsymbol{x}_*)\boldsymbol{m}_0,\ \boldsymbol{\phi}^\top(\boldsymbol{x}_*)\boldsymbol{S}_0\boldsymbol{\phi}(\boldsymbol{x}_*) + \sigma^2\right)$$

9.3.3 Distribuição Posterior



9.3.3 Distribuição posterior

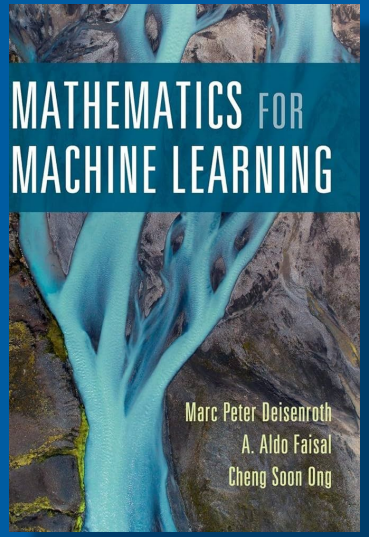
Dado um conjunto de treinamento, a posteriori sobre os parâmetros é:

$$p(\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}) = \frac{p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathcal{Y} \mid \mathcal{X})}$$

Com uma priori Gaussiana conjugada, a posteriori é:

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}) &= \mathcal{N}\left(\boldsymbol{\theta} \mid \boldsymbol{m}_N, \boldsymbol{S}_N\right), \\ \boldsymbol{S}_N &= (\boldsymbol{S}_0^{-1} + \sigma^{-2} \boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1}, \\ \boldsymbol{m}_N &= \boldsymbol{S}_N (\boldsymbol{S}_0^{-1} \boldsymbol{m}_0 + \sigma^{-2} \boldsymbol{\Phi}^\top \boldsymbol{y}) \end{aligned}$$

9.3.4 Previsões Posteriores



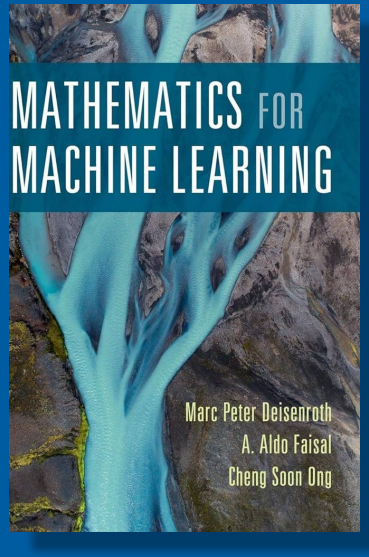
9.3.4 Previsões posteriores

Para prever valores sem ruído, usamos a média e a variância da distribuição posterior:

$$
\begin{aligned}
\mathbb{E}[f(\boldsymbol{x}_*) \mid \mathcal{X}, \mathcal{Y}] &= \mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\phi}^\top(\boldsymbol{x}_*)\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}] = \boldsymbol{\phi}^\top(\boldsymbol{x}_*)\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}] \\
&= \boldsymbol{\phi}^\top(\boldsymbol{x}_*)\boldsymbol{m}_N = \boldsymbol{m}_N^\top \boldsymbol{\phi}(\boldsymbol{x}_*)\,, \\
\mathbb{V}_{\boldsymbol{\theta}}[f(\boldsymbol{x}_*) \mid \mathcal{X}, \mathcal{Y}] &= \mathbb{V}_{\boldsymbol{\theta}}[\boldsymbol{\phi}^\top(\boldsymbol{x}_*)\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}] \\
&= \boldsymbol{\phi}^\top(\boldsymbol{x}_*)\mathbb{V}_{\boldsymbol{\theta}}[\boldsymbol{\theta} \mid \mathcal{X}, \mathcal{Y}]\boldsymbol{\phi}(\boldsymbol{x}_*) \\
&= \boldsymbol{\phi}^\top(\boldsymbol{x}_*)\boldsymbol{S}_N\boldsymbol{\phi}(\boldsymbol{x}_*)\,.
\end{aligned}
$$

A média preditiva coincide com a previsão usando a estimativa MAP, e a variância preditiva reflete a incerteza posterior dos parâmetros.

9.3.5 Calculando a Probabilidade Marginal



9.3.5 Calculando a Probabilidade Marginal

A probabilidade marginal é importante para a seleção de modelos Bayesianos. A probabilidade marginal é dada por:

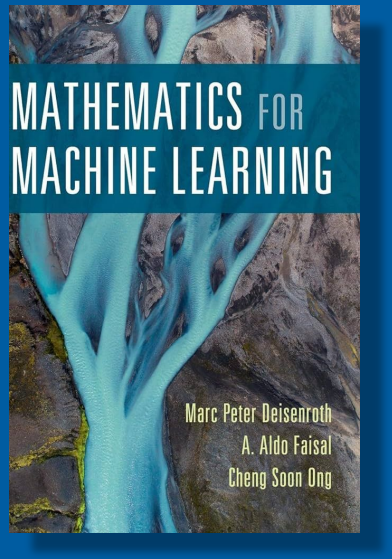
$$p(\mathcal{Y} \mid \mathcal{X}) = \int p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$$

Com a priori Gaussiana conjugada, a probabilidade marginal é Gaussiana. A média e a variância da probabilidade marginal são:

$$\mathbb{E}[\mathcal{Y} \mid \mathcal{X}] = \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{\epsilon}}[\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}] = \boldsymbol{X}\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta}] = \boldsymbol{X}\boldsymbol{m}_0 \,.$$

$$\begin{aligned} \mathrm{Cov}[\mathcal{Y}|\mathcal{X}] &= \mathrm{Cov}_{\boldsymbol{\theta}, \boldsymbol{\epsilon}}[\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}] = \mathrm{Cov}_{\boldsymbol{\theta}}[\boldsymbol{X}\boldsymbol{\theta}] + \sigma^2 \boldsymbol{I} \\ &= \boldsymbol{X}\,\mathrm{Cov}_{\boldsymbol{\theta}}[\boldsymbol{\theta}]\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I} = \boldsymbol{X}\boldsymbol{S}_0\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I} \end{aligned}$$

9.4 Máxima Probabilidade como Projeção Ortogonal



9.4 Máxima Probabilidade como Projeção Ortogonal

A interpretação da estimativa de máxima verossimilhança na regressão linear como uma projeção ortogonal oferece uma compreensão geométrica poderosa do problema de ajuste de linha.

Fundamentos da Regressão Linear em Notação Matricial

O modelo de regressão linear pode ser escrito como:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Onde:

- **y** é um vetor *n × 1* das observações da variável dependente.
- **X** é uma matriz *n × (p + 1)* das observações das variáveis independentes, incluindo uma coluna de 1s para o intercepto.
- **β** é um vetor *(p + 1) × 1* dos coeficientes do modelo.
- **ϵ** é um vetor *n × 1* dos termos de erro.

9.4 Máxima Probabilidade como Projeção Ortogonal

Interpretação Geométrica: Projeção Ortogonal

A interpretação geométrica considera os vetores no espaço euclidiano $\mathbb{R}^n$. A matriz **X** define um subespaço **S** de $\mathbb{R}^n$. O vetor **y** pode ser decomposto em duas componentes ortogonais: uma componente $\hat{\mathbf{y}}$ no subespaço **S** e uma componente de erro $\boldsymbol{\epsilon}$ ortogonal a **S**:

$$\mathbf{y} = \hat{\mathbf{y}} + \epsilon$$

Onde $\hat{\mathbf{y}}$ é a projeção ortogonal de **y** no subespaço **S** gerado por **X**:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$$

9.4 Máxima Probabilidade como Projeção Ortogonal

Projeção Ortogonal

A projeção ortogonal **P** do vetor **y** no subespaço gerado por **X** é dada por:

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

Portanto:

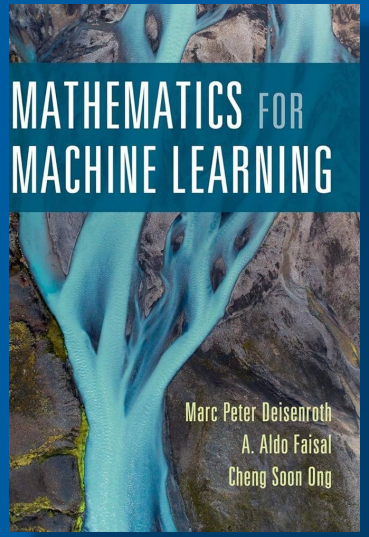
$$\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$$

Propriedades da Projeção Ortogonal

Algumas propriedades importantes da projeção ortogonal incluem:

- **P** é uma matriz idempotente: **PP** = **P**.
- **P** é uma matriz simétrica: $\mathbf{P}^T = \mathbf{P}$.
- O vetor de resíduos $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ é ortogonal ao subespaço gerado por **X**.

Implementação prática



Conclusão

Obrigado!