1^a Prova - FECD B - 2013/02

Renato Assunção

Considere o modelo de regressão linear usual: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ com $\mathbf{X}$ sendo uma matriz $n \times (p+1)$ de posto $p+1$ e com sua primeira coluna sendo todas de 1's. Supomos que $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

O estimador de míminos quadrados de $\boldsymbol{\beta}$ é dado por $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ e os valores preditos estão no vetor $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$. A matrix $\mathbf{H}$ é a matriz de projeção ortogonal de um vetor do $\mathbb{R}^n$ no sub-espaço vetorial formado pelas combinações lineares das colunas de $\mathbf{X}$.

1. Mostre que $\mathbf{H} = \mathbf{H}^2$ e que $\mathbf{H}$ é simétrica.
2. Mostre que $\mathbf{H}\mathbf{X} = \mathbf{X}$ e que $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$, a matrix nula de dimensão $n \times (p+1)$.
3. O vetor de resíduos é $\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}}$. Mostre que o vetor de resíduos é uma transformação linear do vetor de errors gaussianos $\boldsymbol{\varepsilon}$. Isto é, que $\mathbf{r} = \mathbf{H}\boldsymbol{\varepsilon}$.
4. A função-objetivo que o método de mínimos quadrados procura minimizar é a soma de quadrados $C(\boldsymbol{\beta}) = \sum_i (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2$ onde $\mathbf{x}_i'$ é a i-ésima linha da matriz $\mathbf{X}$. Sabemos que RSS é o comprimento ao quadrado de um certo vetor. Que vetor é este?
5. A função $C(\boldsymbol{\beta})$ é minimizada com $\hat{\boldsymbol{\beta}}$ e o seu valor mínimo é dado por $RSS = C(\hat{\boldsymbol{\beta}}) = \sum_i (y_i - \hat{y}_i)^2$. Mostre que $RSS = \boldsymbol{\varepsilon}'\mathbf{H}\boldsymbol{\varepsilon}$. DICA: Use (1) e (3).
6. Na regressão linear com regularização L^2 (ridge regression), o vetor gradiente da função de custo $C(\boldsymbol{\beta})$ que queremos minimizar é igual a
$$\frac{\partial C(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\mathbf{X}'\mathbf{Y} + (\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + \alpha\boldsymbol{\beta}$$
Iguale o gradiente ao vetor zero e resolva a equação para encontrar a solução $\hat{\boldsymbol{\beta}}$ da regressão ridge.
7. Seja $f(w_1, w_2, w_3) = w_1 w_2 + \exp(-w_2^2) + w_1 w_3$. Obtenha o seu vetor gradiente ∇f.
8. Responda V ou F:
   (a) Inserir uma nova feature-coluna na matriz $\mathbf{X}$ nunca diminui e, em geral, aumenta o valor do R^2.
   (b) Mesmo quando o modelo de regressão linear é o modelo que verdadeiramente gera os dados, teremos $\boldsymbol{\beta} \neq \hat{\boldsymbol{\beta}}$.
   (c) $\hat{\boldsymbol{\beta}}$ é um vetor aleatório com distribuição gaussiana.
   (d) Para regularização ridge no modelo de regressão, minimizamos em $\boldsymbol{\beta}$ a função de custo $||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2 + \alpha||\mathbf{w}||^2$ onde $\boldsymbol{\beta}' = (b, \mathbf{w})'$.
9. Suponha que X_1, X_2, X_3 seja uma amostra aleatória de v.a.'s i.i.d. com a distribuição Poisson com parâmetro $\lambda > 0$ desconhecido. Isto é, $\mathbb{P}(X_i = k) = \lambda^k e^{-\lambda}/k!$ com $k = 0, 1, 2 \ldots$. Observamos a seguinte realização dessas variáveis numa amostra específica: $x_1 = 1, x_2 = 0, x_3 = 1$. Obtenha o MLE de λ. (OBS: $0! = 1$).
10. Suponha que $X_1, \ldots, X_n$ forme uma amostra aleatória de v.a.'s i.i.d. com a seguintes densidade de probabilidade (caso contínuo): $f(x; \theta) = \theta e^{-\theta x}$ para $x > 0$ com $\theta \in (0, \infty)$ (densidade exponencial, contínua). Encontre o MLE de θ.