Aprendizagem de Máquina I – Lista 06
Referente aos slides "08 Classificacao e classificadores gaussianos"

Prof. Hugo Carvalho

Fontes para os exercícios

- [ITSL] Gareth James, Daniela Witten, Trevor Hastie, Rob Tibshirani & Jonathan Taylor - *An Introduction to Statistical Learning, with Applications in Python* [baixe aqui]
- [AME] Rafael Izbicki & Tiago Mendonça dos Santos - *Aprendizado de Máquina: Uma Abordagem Estatística* [baixe aqui]

Questões do livro

- [ITSL] Capítulo 4
  - Questões conceituais: 2, 3, 5, 10, 11

Questões avulsas

Questão 1: O objetivo dessa questão é provar o resultado abaixo:

Teorema 1. *Seja X uma variável aleatória contínua com função densidade de probabilidade $p(x)$, média μ, e variância σ^2, sendo ambos os valores finitos. Então, dentre todas as distribuições com média μ e variância σ^2, a distribuição normal, com densidade*

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right)$$

é a única que maximiza a entropia diferencial*, dada por*

$$\mathcal{H}(p) = -\int_{-\infty}^{\infty} p(x) \log p(x)\, dx.$$

Para isso, siga o roteiro abaixo.

a) Familiarize-se com as duas quantidades e interpretações abaixo:

   - **Entropia diferencial**, $\mathcal{H}(p)$: Quantifica o nível médio de incerteza ou informação associado aos possíveis estados ou resultados da variável. Isso mede a quantidade esperada de informação necessária para descrever o estado da variável, considerando a distribuição de probabilidades entre todos os estados possíveis.
   - **Divergência de Kullback-Leibler**, $D_{\mathrm{KL}}(p \,\|\, q)$: Mede quão diferente uma distribuição de probabilidade aproximada q é de uma distribuição de probabilidade verdadeira p. Uma outra interpretação da divergência de KL de p em relação a q é o excesso esperado de surpresa ao usar

a aproximação q em vez de p, quando a distribuição real é p. A divergência de Kullback-Leibler é calculada como:

$$D_{\mathrm{KL}}(p \,\|\, q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx.$$

A quantidade $D_{\mathrm{KL}}(p \,\|\, q)$ é sempre maior que ou igual à zero, sendo zero se e somente se $p = q$ em quase todo ponto.

b) Partindo da divergência KL e utilizando o fato de que tal quantidade é sempre maior que ou igual à zero, mostre que

$$\mathcal{H}(p) \leq -\int p(x) \log q(x)\, dx.$$

c) Utilizando que $\mathbb{E}[X] = \mu$ e que $\mathbb{V}(X) = \sigma^2$, substitua a expressão para a densidade da normal em q no termo $\int p(x) \log q(x)\, dx$ da desigualdade acima e mostre que

$$\int p(x) \log q(x)\, dx = \frac{1}{2} \log(2\pi e \sigma^2).$$

d) Juntando a desigualdade do item b) com o calculado no item c), conclua que

$$\mathcal{H}(p) \leq \frac{1}{2} \log(2\pi e \sigma^2).$$

e) Mostre que a entropia da distribuição normal de média μ e variância σ^2 é igual ao valor do lado direito da desigualdade acima.

f) Conclua o resultado desejado.

Questão 2: Generalize e prove o resultado da Questão 1 para distribuições normais multivariadas.