

Lista 2 – entregar os exercícios 5, 7, 10 e 11 até 17/10

O objetivo principal desta lista é propiciar uma “motivação” para a revisão de alguns tópicos vistos até agora.

1. Calcule a matriz de covariância amostral para os dados do arquivo `lista2_ex1.dat`. Cada linha do arquivo corresponde a uma realização de 4 variáveis.

Apesar de existirem várias bibliotecas com funções que realizam exatamente o cálculo acima, para entender exatamente do que se trata é importante realizar o cálculo seguindo a definição. Se achar conveniente, escreva um programa que realiza o cálculo e depois compare com o resultado de algum pacote.

2. (Classificador de Bayes) Script em R `classificadorBayes_exemplo.R`

Script que ajuda a visualizar o limiar de decisão no caso de duas classes, ambas com fdp gaussiana. Fica claro o efeito da probabilidade a priori das classes (0.1 e 0.9, respectivamente)

Para executar o script, basta executar o R e então digitar

```
source('classificadorBayes_exemplo.R')
```

3. (Classificador de Bayes e Classificador Naive de Bayes) Há alguns softwares didáticos interessantes nas páginas de Christian Borgelt (<http://www.borgelt.net/software.html>). Entre eles,

- BayesGUI (<http://www.borgelt.net/bcgui.html>), para a indução (treinamento) de classificadores de Bayes e naive de Bayes.

Utilize os arquivos disponibilizados (`iris.dom/iris.tab` e `ex.dom/ex.trn.tab/ex tst.tab/ex all.tab`) para a indução de classificadores full bayes e naive bayes.

- BCView (<http://www.borgelt.net/bcview.html>), para a visualização do classificador de Bayes.

Visualize a região correspondente às três classes, determinadas pelos classificadores obtidos acima, usando o BCView.

4. (Estimação não paramétrica de fdp) `density_estimation.R`

Script que ajuda a experimentar a estimação da fdp a partir de amostras de uma gaussiana. Experimentar diferentes quantidades de amostras e tipos e larguras de janelas é interessante.

Para executar o script, basta executar o R e então digitar

```
source('density_estimation.R')
```

5. (Perceptron training) Refaça a simulação feita em sala de aula para determinar a reta que separa um conjunto de amostras. Considere o mesmo conjunto de amostras, i.e., $\{\mathbf{x}_1 = (0, 0)^t, \mathbf{x}_2 = (0, 1)^t, \mathbf{x}_3 = (1, 0)^t, \mathbf{x}_4 = (1, 1)^t\}$ (no qual os dois primeiros são da classe 1 e os dois últimos da classe 2) e a função de custo $J_p(\mathbf{a}) = \sum_{\mathbf{y} \in \mathcal{Y}(\mathbf{a})} (-\mathbf{a}^t \mathbf{y})$, porém o algoritmo batch para minimização de J_p .

Algorithm 1: Algoritmo perceptron (batch)

Entrada: conjunto de exemplos (estendidos) $\{\mathbf{y}\}$, η e θ

Saída : vetor de pesos \mathbf{a}

$\mathbf{a} \leftarrow \mathbf{0}$;

$k \leftarrow 0$;

repeat

$k \leftarrow k + 1$;

$\Delta = \eta \sum_{\mathbf{y} \in \mathcal{Y}(k)} \mathbf{y}$; // $\mathcal{Y}(k)$ é o conjunto de amostras com classificação errada

$\mathbf{a} = \mathbf{a} + \Delta$;

until $|\Delta| < \theta$;

Qual o efeito do uso de diferentes valores para η e θ ?

6. (Perceptron training demo) Experimente a simulação do treinamento de um perceptron. Por exemplo, o Java applet em <http://eecs.wsu.edu/~cook/ai/lectures/applets/perceptron/>.
7. (PCA) Seja a matriz de covariância

$$\Sigma = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix}$$

e a matriz de correlação correspondente

$$\rho = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

Calcule os componentes principais (PCA) para ambas as matrizes. Discuta os resultados (em particular, se houverem diferenças, explique o significado delas).

8. (Discriminante linear de Fisher) Conforme vimos, no caso do discriminante linear de Fisher, deseja-se encontrar $\mathbf{a} \in \mathbb{R}^d$, de forma que a transformação linear $y = \mathbf{a}^t \mathbf{x}$ seja a que maximiza

$$sep = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma_y}$$

na qual \bar{y}_1 , \bar{y}_2 e σ_y são, respectivamente, a transformação do ponto médio da classe 1, a transformação do ponto médio da classe 2 e o desvio padrão dos pontos transformados. Portanto, a formulação de Fisher considera que a matriz de covariância das duas classes são iguais.

Supondo que ambas as classes possuem distribuição normal, com a mesma matriz de covariância, digamos Σ , qual será o discriminante linear de Fisher? Existe relação desse discriminante com a superfície de decisão relativa ao classificador ótimo de Bayes?

9. (Multilayer perceptron training demo) Experimente o treinamento de uma MLP. Por exemplo, o applet em <http://lcn.epfl.ch/tutorial/english/general/html/index.html>.

10. (Avaliação e comparação de classificadores)

Leia o artigo

- Steven L. Salzberg, On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach, Data Min. Knowl. Discov., 3, pp.317-328, 1993

Comente um ou mais pontos do artigo que você considerou importantes e também os pontos que você não entendeu claramente.

Outro artigo, também interessante, é:

- Thomas G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 10:1895-1923, 1998.

11. Comente sobre os demais exercícios, caso você tenha tentado resolver ou testar.

PS.: os arquivos com os scripts, dados e artigos estão todos no PACA