

Aprendizado de Máquina e Reconhecimento de Padrões (UTFPR/CPGEI) - Lista de Exercícios 3

Tópicos: Classificadores Lineares.

1. No espaço 2-D, têm-se duas classes equiprováveis com médias $\mathbf{m}_1 = [0, 0]^T$ e $\mathbf{m}_2 = [1, 5, 1, 5]^T$ e matrizes de covariância $\mathbf{S}_1 = \mathbf{S}_2 = 0,2\mathbf{I}$, sendo \mathbf{I} uma matriz identidade 2×2 .
 - a) Gere e plote um conjunto de dados X_1 , contendo 200 exemplos por classe (400 no total), a ser utilizado como conjunto de treinamento (utilize o valor 10 como 'seed' e a função randn Matlab). Gere um segundo conjunto de dados X_2 contendo 200 exemplos por classe para ser utilizado como conjunto de teste (utilize o valor 100 como 'seed' e a função randn Matlab).
 - b) Com base em X_1 , gere seis classificadores do tipo SVM para separar as duas classes, variando $C = 0,1, 0,2, 0,5, 1, 2, 20$. Utilize o algoritmo de Platt (SMO2.m) e $tol = 0,001$.
 - I. Calcule os erros de classificação dos conjuntos de treino e teste.
 - II. Conte o número de vetores suporte.
 - III. Calcule o valor da margem ($2/||\mathbf{w}||$).
 - IV. Plote o classificador em conjunto com as margens. Comente os resultados obtidos nas etapas anteriores. Qual a influência do parâmetro C ?
2. Considere um problema de duas classes, bidimensional, com os seguintes parâmetros das classes (distribuição Gaussiana):

$$\begin{aligned}\mu_1 &= [0, 2]^T \\ \Sigma_1 &= \begin{bmatrix} 4 & 1.8 \\ 1.8 & 1 \end{bmatrix} \\ \mu_2 &= [0, 0]^T \\ \Sigma_2 &= \begin{bmatrix} 4 & 1.8 \\ 1.8 & 1 \end{bmatrix}\end{aligned}$$

- i. Gere os dados das duas classes e plote os resultados, considerando 1500 exemplos por classe e um conjunto \mathbf{X} para treinamento e \mathbf{X}_{test} para teste.
- ii. Classifique os exemplos de \mathbf{X}_{test} utilizando a regra de decisão Bayesiana.
- iii. Crie um modelo com base em Logistic Regression Classifier utilizando o conjunto \mathbf{X} e avalie o desempenho desse classificador no conjunto \mathbf{X}_{test} .
- iv. Comente e compare os resultados obtidos em ii e iii.
- v. Repita e compare os passos de i a iv, considerando agora:

$$\Sigma_2 = \begin{bmatrix} 4 & -1.8 \\ -1.8 & 1 \end{bmatrix}$$

- vi. Repita a análise para a Logistic Regression, considerando agora regularização com diferentes valores de lambda (definidos por você). Qual o efeito da regularização, comparando o desempenho de treinamento e teste?