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

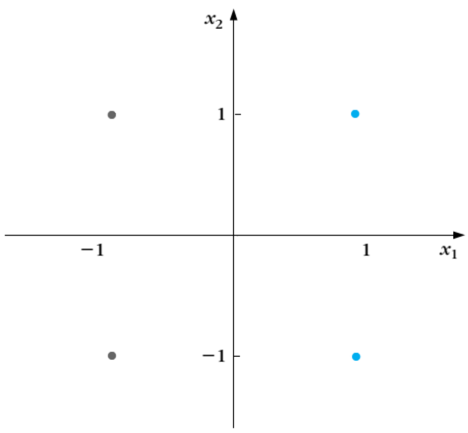
**Tópicos**: Classificadores Lineares.

## Parte 1 – Exercícios Teóricos:

1. Deduza a equação matricial (equações normais) do classificador baseado em mínimos quadrados.

2. Como uma SVM pode ser modelada no caso de um problema multi-classes (número de classes maior ou igual a três)? Quais as abordagens mais usuais?

3. Dados os exemplos de treinamento abaixo, obtenha a linha de separação ótima (classificador SVM linear), utilizando as condições KKT, caso necessário. Restrinja a busca pelo ótimo nas linhas que cruzam a origem.




## Parte 2 – Exercícios Práticos:

1. Gere um conjunto com $N_1$=1000 exemplos, sendo que os 500 primeiros pertencem à classe $\omega_1$, definida por uma distribuição Gaussiana com média $m_1 = [0, 0, 0, 0, 0]^T$ e o restante à classe $\omega_2$ com $m_2 = [2, 2, 0, 2, 2]^T$. Ambas as classes têm a seguinte matriz de covariância:

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 10^{-350} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Cada vetor de dados desse conjunto é aumentado por uma sexta coordenada, que vale +1 para todos os vetores. Com isso, $X_1$ é definido por uma matriz $(l + 1) \times N$ , sendo suas colunas os exemplos desse conjunto de dados (para reprodutibilidade de resultados, utilize ‘seed’ = 0 na função randn do Matlab). Adicionalmente, gere um conjunto $X_2$ com 10.000 exemplos, utilizando a mesma distribuição utilizada para $X_1$ (para reprodutibilidade de resultados, utilize ‘seed’ = 100 na função randn do Matlab).

a) Utilize o método de mínimos quadrados (sem regularização) para estimar os parâmetros. Calcule também o fator $X_1X_1^T$. O que ele significa?

b) Repita o exercício anterior, para a versão regularizada do mínimos quadrados para o parâmetro de regularização igual a 0,1.

c) Comente os resultados obtidos nos passos anteriores. Qual o efeito da regularização, comparando o desempenho de treinamento e teste?

d) Calcule os erros de classificação obtidos nas etapas 1 e 2, para o conjunto de dados $X_2$.

2. Considere um problema de três classes com classes equiprováveis $\omega_1$, $\omega_2$ e $\omega_3$. As classes são modeladas com distribuições Gaussianas com $\mathbf{m}_1 = [0,0,0]^T$, $\mathbf{m}_2 = [1,2,2]^T$ e $\mathbf{m}_3 = [3,3,4]^T$, respectivamente. Todas as distribuições têm a mesma matriz de covariância:

$$S = \begin{bmatrix} 0.8 & 0.2 & 0.1 \\ 0.2 & 0.8 & 0.2 \\ 0.1 & 0.2 & 0.8 \end{bmatrix}$$

a) Gere um conjunto de dados de treinamento ($\mathbf{X}_1$) e teste ($\mathbf{X}_2$) com 1000 e 10000 exemplos, respectivamente. Determine o conjunto de parâmetros $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ utilizando o método de mínimos quadrados nos dados de $\mathbf{X}_1$. Utilize os dados de $\mathbf{X}_2$ para determinar os erros de classificação por classe e global.

b) No modelo de mínimos quadrados, quando utiliza-se 0 e 1 para modelar a resposta desejada das classes (rótulos das classes), o modelo retorna como saída a probabilidade a posteriori. Isto é, se $\mathbf{w}_j$ é a estimativa de parâmetros do j-ésimo classificador, então:

$$g_j(x) \equiv w_j^T x \approx P(\omega_j|x)$$

Comprove isso utilizando como base o conjunto $\mathbf{X}_2$.

c) Calcule o erro de classificação do classificador Bayesiano (ótimo) e compare com os resultados obtidos para o classificador baseado em mínimos quadrados do item “a”.

3. 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 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 50 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 ($2/||\boldsymbol{w}||$).
   IV. Plote o classificador em conjunto com as margens. Comente os resultados obtidos nas etapas anteriores. Qual a influência do parâmetro C?

4. Considere um problema de duas classes, bidimensional, com os seguintes parâmetros das classes (distribuição Gaussiana):

$$\boldsymbol{\mu}_1 = [0,\ 2]^{\mathrm{T}}$$

$$\Sigma_1 = \begin{bmatrix} 4 & 1.8 \\ 1.8 & 1 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = [0,\ 0]^{\mathrm{T}}$$

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

   i. Gere os dados das duas classes e plote os resultados, considerando 1500 exemplos por classe e um conjunto **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 utilizando o conjunto **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?