



Instituto de Computação
Unicamp



MO 416 - Introdução à Inteligência Artificial

1º Semestre de 2010 - Lista 3

Prof. Siome Goldenstein

Entrega: Quinta, 24/06/2010, **no início da aula.**

Questão 1 - Metodos Lineares

1. **Geração de dados sintéticos.** Crie um dataset sintético em 2D com duas classes:
 - (a) A classe A possui 100 amostras de uma distribuição Gaussiana com $\mu = [-2, -2]^T$ e com $\Lambda = I$.
 - (b) A classe B possui 50 amostras de uma distribuição Gaussiana com $\mu = [-1, -1.5]^T$ e com $\Lambda = I$ e 50 amostras de uma distribuição Gaussiana com $\mu = [1.5, -1]^T$ e com $\Lambda = I$.
2. Implemente a regressão linear e o LDA para classificação dos seus dados, e desenhe a superfície (reta) de divisão.
3. Mapeie o problema de $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ da seguinte forma $[x, y] \rightarrow [x^2, y^2, xy]$, e aplique os dois algoritmos de classificação nos dados. Desenhe as superfícies de decisão (que agora não serão mais retas).

Questão 2 - Redes Bayesianas

Calcule $P(t_1|z_1)$ no Exemplo 3.2 do Livro “Learning Bayesian Networks” do Neapolitan (página128).

Questão 3 - Classificadores

O conjunto de dados que utilizaremos possui 18828 mensagens de texto, de tamanho variado, coletadas de 20 newsgroups. Os dados necessários para esta questão estão disponíveis em:

<http://www.ic.unicamp.br/~siome/teaching/2010/mo416-0110/data/news-groups-20.zip>

1. Crie cinco partições balanceadas dos dados para a validação cruzada a ser usada na avaliação de classificadores.
2. Utilize diferentes formas de mapear os dados em vetores de características e utilize implementações prontas de árvores de decisão e naïve Bayes para fazer a classificação dos dados (não utilize Weka). Reporte seus resultados sob a forma de matrizes de confusão com médias e desvios padrões.
3. Utilize métodos OvO e OvA de coletâneas de classificadores LDA e SVM para fazer a classificação dos dados. No caso do SVM, explore diferentes opções de Kernel (como kernel dedicados para strings, por exemplo).