P2 – Tópicos em Aprendizagem de Máquina e Classificação de Padrões – 19/11/2010

1-Seja $p(x) \sim U(0, a)$ uma distribuição uniforme de 0 até a, e seja uma janela de Parzen definida por $\varphi(x) = e^{-x}$ para $x > 0$ e 0 para $x \leq 0$.

(a) mostrar que média da estimativa desta janela de Parzen é

$$\bar{p}_n = \begin{cases} 0 & x < 0 \\ \frac{1}{a}(1 - e^{-x/h_n}) & 0 \leq x \leq a \\ \frac{1}{a}(e^{a/h_n} - 1)e^{-x/h_n} & a \leq x \end{cases}$$

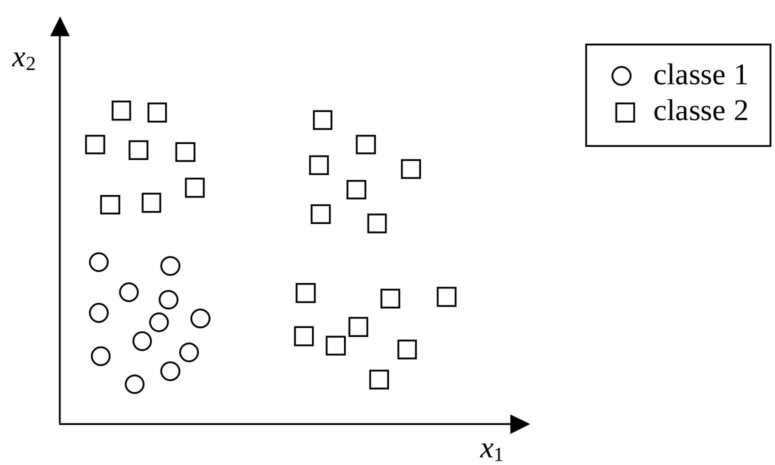
(b) esboçar $\bar{p}_n(x)$ versus x para $a = 1$, $h_n = 1$, 1/4 e 1/16

(c) quão pequeno dever ser o valor de h_n para se obter menos que 1% de *bias* em 99% do intervalo $0 < x < a$?

(d) determinar h_n para esta condição se $a = 1$ e esboçar $\bar{p}_n(x)$ no intervalo $0 \leq x \leq a$.

2-Supor um problema de classificação com c categorias onde amostramos a distribuição $p(\mathbf{x})$ e subseqüentemente treinamos uma rede probabilística (PNN) com o Algoritmo 1. Mostrar que mesmo que as distribuições a priori $P(\omega_i)$ sejam diferentes (o número de pontos em cada categoria são diferentes) o método de reconhecimento considera apropriadamente as distribuições a priori.

3-Considerar a figura abaixo, com pontos em um espaço bidimensional pertencentes a duas classes. Propor uma base de regras *fuzzy* lingüísticas para que defina um classificador apropriado. Definir os universos, a granularização dos universos, escolher funções de pertinência apropriadas nos universos correspondentes. Sugerir também uma semântica para as regras *fuzzy* e um procedimento de inferência para o classificador.



4-Expandir o lado esquerdo da expressão abaixo para obter o lado direito, o qual expressa o erro quadrático médio como a soma do quadrado da polarização (*bias*^{2}) com a variância. Pode a polarização ter um valor negativo? E a variância?

$$E_D[(g(\mathbf{x};D)-F(\mathbf{x}))^2] = (E_D[g(\mathbf{x};D)-F(\mathbf{x})])^2 + E_D[(g(\mathbf{x};D)-E_D[g(\mathbf{x};D)])^2]$$

5-Mostrar que a expressão para a média $\mu_{(.)}$ das médias *leave one out* $\mu_{(i)}$ é equivalente à média amostral $\hat{\mu}$, isto é, mostrar que $\mu_{(.)} = \hat{\mu}$.