

CC-226 Aula - Lista de Exercícios

Carlos Henrique Q. Forster - Instituto Tecnológico de Aeronáutica

2008

1 Lista de Exercícios

1. Explicar o programa classificador linear de peixes da primeira aula.
2. Provar que aplicando-se a inversa de uma distribuição cumulativa de probabilidade F^{-1} a uma variável aleatória uniformemente distribuída U entre 0 e 1, obtemos uma variável aleatória $Y = F^{-1}(u)$ com a distribuição de probabilidades correspondente $f(y) = F'(y)$.
3. Encontrar um limite superior para a probabilidade de se obter um valor maior que k vezes a média para uma distribuição qualquer de uma variável aleatória com valores não-negativos.
4. Descrever em detalhes o algoritmo de k-vizinhos mais próximos para classificação.
5. Descrever em detalhes o algoritmo de janela de Parzen para classificação e regressão.
6. Construir um exemplo ilustrativo de rede bayesiana na forma de grafo acíclico com pelo menos 6 nós. Explicar sua construção e a escolha dos valores de probabilidades condicionais.
7. Descrever em detalhes o algoritmo naïve Bayes para classificação.
8. (DHS 2.4 ex 13) Em problemas de classificação de padrões, pode haver a opção de associar um padrão a uma das c classes ou rejeitar o padrão como irreconhecível. Seja a perda dada por:

$$\lambda_{ij} = \left\{ \begin{array}{ll} 0 & i = j \\ \lambda_r & i = c + 1 \\ \lambda_s & \text{caso contrário} \end{array} \right\}$$

Onde λ_r é a perda no caso de rejeição e λ_s é a perda no caso de classificações incorretas quaisquer. Mostre que o risco mínimo é obtido se decidirmos pela classe C_i se $P(C_i|x) \geq 1 - \lambda_r/\lambda_s$ ou por rejeição do padrão caso contrário. O que acontece se $\lambda_r = 0$? O que acontece se $\lambda_r > \lambda_s$?

2 Implementações

1. Histograma com área correspondente à densidade de probabilidade. Ver especificação nos slides da primeira aula.
2. Análise de regressão múltipla. Encontrar uma expressão para a estimação da porcentagem de gordura no corpo. Utilizar validação cruzada e erro RMS para medir se a expressão é uma boa generalização.
3. Regressão não-linear utilizando algum dos métodos (RBF, Janela de Parzen ou Vizinhos Mais Próximos) para a base BOSTON. Estimar Poluição (NOX) e valor da habitação (MEDV) em função de latitude e longitude. Mostrar gráfico com curvas de nível. Utilizar validação cruzada e erro RMS para avaliar.
4. Construir dois classificadores diferentes para a base WINE utilizando pelo menos dois dos métodos aprendidos (k-nn, kernel, hiperplano LMS, naïve bayes, classificador bayesiano, rede bayesiana) e comparar os resultados. Utilizar validação cruzada e taxa de acerto.

3 Prazos e Valores

- Os exercícios devem ser entregues em papel, na aula do dia 30 de abril (depois da semaninha).
- Para essa aula, os alunos devem vir preparados para apresentar as implementações, seus resultados e a lista de exercícios. Os alunos devem ter a iniciativa de apresentar sua solução para discussão na classe e a participação na aula (seja apresentando ou discutindo) será objeto de avaliação.
- Não haverá prova. A lista de exercícios corresponderá a 40% da nota do bimestre e as implementações a 60%.
- Os alunos devem se organizar em grupos e escolher um tema de trabalho até a aula do dia 16 de abril. Não haverá aula dia 9 de abril a fim de que os alunos utilizem esse tempo para se organizar. Todo tema de trabalho deve apontar um artigo científico de importância, seja sobre uma técnica ou sobre uma aplicação.
- As apresentações finais devem ser agendadas para as semanas 13 a 16. Correspondo aos dias: 28/maio 4/junho 11/junho e 18/junho. Uma proposta é sortear o grupo para cada data da primeira a última. Grupos maiores terão probabilidade maior de sorteio em uma semana inicial.
- Todos membros do grupo devem apresentar uma parte do trabalho. A dificuldade do assunto escolhido pelo participante individual será considerada na avaliação.
- O trabalho final deve ser entregue de forma completa (dados, programa, relatório, slides) em um CD.