B529: Homework 4

Nathan Byers

Wednesday, April 21, 2015

Question 1

For 2-dimensional data points, we define a kernel function $K(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^T x')^3$, what is its corresponding transformation function $z = \Phi(x)$. (15 points)

Answer 1

$$K(\mathbf{x}', \mathbf{x}) = (1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}')^3 = \left(1 + \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}\right)^3$$

$$= (1 + x_1x'_1 + x_2x'_2)(1 + x_1x'_1 + x_2x'_2)(1 + x_1x'_1 + x_2x'_2)$$

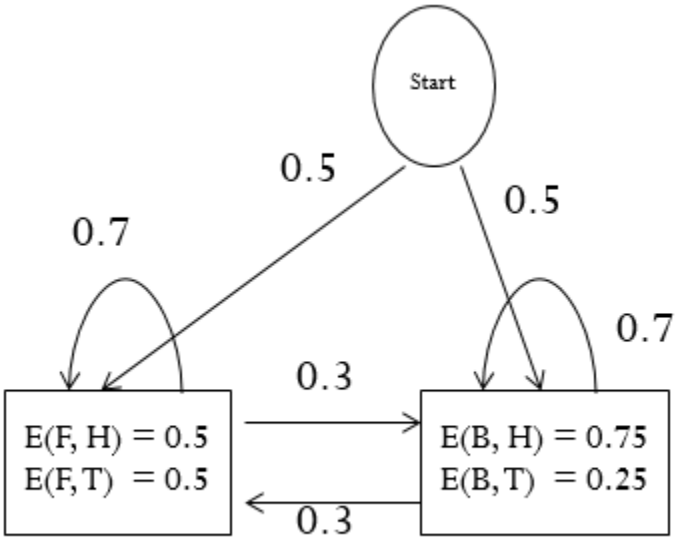
$$= (1 + 2x_1x'_1 + 2x_2x'_2 + x_1^2x'^2_1 + 2x_1x'_1x_2x'_2 + x_2^2x'^2_2)(1 + x_1x'_1 + x_2x'_2)$$

$$= 1 + 3x_1x'_1 + 3x_2x'_2 + 3x_1^2x'^2_1 + 6x_1x'_1x_2x'_2 + 3x_2^2x'^2_2 + x_1^3x'^3_1 + 3x_1^2x'^2_1x_2x'_2 + 3x_1x'_1x_2^2x'^2_2 + x_2^3x'^3_2$$

$$\Phi(\mathbf{x})^{\mathrm{T}}\Phi(\mathbf{x}) = \begin{bmatrix} 1 & \sqrt{3}x_1 & \sqrt{3}x_2 & \sqrt{3}x_1^2 & \sqrt{6}x_1x_2 & \sqrt{3}x_2^2 & x_1^3 & \sqrt{3}x_1^2x_2 & \sqrt{3}x_1x_2^2 & x_2^3 \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{3}x'_1 \\ \sqrt{3}x'_2 \\ \sqrt{3}x'^2_1 \\ \sqrt{6}x'_1x'_2 \\ \sqrt{3}x'^2_2 \\ x'^3_1 \\ \sqrt{3}x'^2_1x'_2 \\ \sqrt{3}x'_1x'^2_2 \\ x'^3_2 \end{bmatrix}$$

Question 2

In the following Hidden Markov model, the observation sequence is TTHH. If the hidden path is BBFF, what is the probability that the sequence is generated by the path BBFF (10 points). If the hidden path is unknown, generate the graph for the decoding problem (15 points). What is the heaviest path in the graph? (10 points)



Answer 2

Question 3

Using the following data set for credit card application, compute the mutual information for each feature (15 points), compute the chi-square value for each feature (15 points)

Input			Output
Age	**Income**	**Gender**	**Risk**
<25	>50K	M	High
<25	>50K	F	High
≥25	<50K	F	High
≥25	>50K	F	Low
≥25	>50K	M	Low
<25	<50K	M	High

Question 4

Given a 2-dimensional data set (0,1) (2,3), (3,4), (2,4), (4,6). Please use R to find two resulting vectors in PCA and plot the data points and the vectors (20 points).

Answer 4