MSPR 5 Probabilities (Due: 11.10.2015, 12 p.m. (noon))

Dr. Hendrik Purwins, Assistant Professor
Jan Stian Banas, TA
Dept. Architecture, Design and Media Technology, Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark

1. (Feedback) Please give us feedback on the last lecture and homework: `http://goo.gl/forms/PjC11i5LM8` Thanks!

2. Implement the 2-dimensioanl Gaussian distribution for a J=2-dimensional variable $\mathbf{x}$ in Matlab (20p)

$$\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma}) = \frac{1}{2\pi^{J/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$$

   Do not use the function `mvnpdf`

3. Load the file `gauss3.mat`. You will find data vectors `A0_tr, A0_tst, A1_tr, A1_tst` with the points in the rows. These data come from two distributions D_0 and D_1. The suffixes *_tr* and *_tst* indicate which data vector should be used for training and testing the classifier. By *training* we mean estimating the true mean μ_0 (μ_1) and the true variance σ_0^2 (σ_1^2) of the two Gaussians, the examples have been drawn from. Use the training data `A0_tr` and `A1_tr` to calculate the two parameter pairs (mean and variance) for the two distributions D_0 and D_1. (20 P)

4. For each vector `A0_tr, A1_tr` plot the histogram in the same figure but with different color. Apply the Matlab function `[h, b]=hist(A)` to each of the training data vectors. `hist` returns the positions on the x-axis (`b`) and the heights of the bars (`h`). Then use the Matlab function `bar` to plot the histograms `h` over `b`. Normalize the height of the histogram bars `h` by dividing it with the distance `b(2) - b(1)` between two bar positions and the number of components in the data vector of that class. Using the calculated parameter pairs of the Gaussians $(\bar{x}_0, s_0^2)$ and $(\bar{x}_1, s_1^2)$ plot the corresponding Gaussian probability distribution over the plot of the histogram of that data vector. For plotting the Gaussian probability distribution, use a spacing of $-10 : 0.1 : 10$ for the argument $\mathbf{x}$. Make sure that histogram/normal distribution of the same class have the same color in the plot. Comment on how well the normal distribution fits the histogram. (40 P)

5. Perform Principal Component Analysis on the iris dataset (including all 4 features Sepal Length/Width and Petal Length/Width. Take the scores for the eigenvectors with highest eigenvalues. Fit two Gaussians to instance 6-50 of the versicolor and the virginica iris species separately and use these Gaussians to predict the iris type of the first 5 instances of versicolor and virginica without using their true labels. Give the number of erroneous predictions. Use

(a) the full covariance matrix (20p)

(b) the covariance matrix of type $\mathbf{S} = \sigma^2\mathbf{I}$, where σ^2 is the mean variance of all 4 features. (20p)

6. Self Assessment: Check the exercises that you have seriously worked on.

2	3	4	5 a	5 b