Learning from Data, Tutorial Sheet for Week 4

School of Informatics, University of Edinburgh

Instructor: Amos Storkey

The Gaussian distribution in one dimension is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

and satisfies $\int_{-\infty}^{\infty} p(x)dx = 1$.

1. Show that $\int_{-\infty}^{\infty} xp(x)dx = \mu$.

2. Show that $\int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \sigma^2$.

3. Consider data $x^i, i = 1, \ldots, P$. Show that the Maximum Likelihood estimator of μ is $\hat{\mu} = \frac{1}{P}\sum_{i=1}^{P} x^i$ and that the ML estimate of σ^2 is $\hat{\sigma^2} = \frac{1}{P}\sum_{i=1}^{P}(x^i - \mu)^2$

4. A training set consists of one dimensional examples from two classes. The training examples from class 1 are $\{0.5, 0.1, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.35, 0.25\}$ and from class 2 are $\{0.9, 0.8, 0.75, 1.0\}$. Fit a (one dimensional) Gaussian using Maximum Likelihood to each of these two classes. Also estimate the class probabilities p_1 and p_2 using Maximum Likelihood. What is the probability that the test point $x = 0.6$ belongs to class 1?

5. Given the distributions $p(x|class1) = N(\mu_1, \sigma_1^2)$ and $p(x|class2) = N(\mu_2, \sigma_2^2)$, with corresponding prior occurrence of classes p_1 and p_2 ($p_1 + p_2 = 1$), calculate the decision boundary explicitly as a function of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p_1, p_2$. How many solutions are there to the decision boundary, and are they all reasonable?