Homework 6: Linear Classification

Solutions to this exercise sheet are to be handed in before the lecture on Friday, 30.11.12. Code (.m-file) and output (.mat-file, figures as .pdf-files) for the matlab-questions should be either submitted by email to `nicolas.ludolph@student.uni-tuebingen.de` (subject: [ML1] Exercise 6) or uploaded to Ilias. Use comments to explain your code. Please adhere to the file naming convention: `Homework6_<YourName>.<ext>`.

1. (15 points) **Linear classification and the logistic function**. This exercise will be concerned with the logistic function $\sigma(s) = 1/(1 + \exp(-s))$ as well as its connection with linear classification in Gaussian models.
   (a) Show that the logistic function satisfies $\sigma(-s) + \sigma(s) = 1$ and find the first two derivatives of $\sigma(s)$, $\sigma'(s)$ and $\sigma''(s)$.
   (b) Plot $\sigma(s)$ as well as $\log(\sigma(s))$ as a function of s (either using matlab or with pen and paper, a rough plot which captures the qualitative features of the functions is sufficient). Explain why, for large $s > 0$, $\log(\sigma(s)) \approx 0$ and $\log(\sigma(-s)) \approx -s$.
   (c) Suppose that we have data from two classes, and the data within each class is Gaussian distributed with the same covariance, i.e. $x|t = 1 \sim \mathcal{N}(\mu_+, \Sigma)$ and $x|t = -1 \sim \mathcal{N}(\mu_-, \Sigma)$, and that the two classes the same prior probabilities $\pi_+ = P(t = +1) = \pi_- = P(t = -1) = 0.5$. Show that the conditional probability of belonging to the positive class can be written as a logistic function $P(t = 1|x) = \sigma(\omega^\top x + \omega_o)$ and identify the corresponding parameters ω and ω_o.

2. (15 points) **Linear Classification [matlab]** Download the file `HomeWork6.mat`, in which you will find training data `xTrain` (a matrix of size $N = 500$ by $D = 2$) with labels `tTrain`. Your job will be to train and compare two classification algorithms on this data.
   (a) Calculate the means and the covariances of each of the two classes, as well as the average covariance $\Sigma = \frac{1}{2}\Sigma_+ + \frac{1}{2}\Sigma_-$. Use μ_+, μ_- and Σ to the weight vector ω and offset ω_o. of the Gaussian linear discriminant analysis used in lectures.
   (b) Plot the data as well as the decision boundary into a 2-D plot, and calculate the (training) error rate of the algorithm, i.e. the proportion of points in the training set which were misclassified by it. Use the data in `xTest` and `tTest` to also calculate its error rate on the test set.
   (c) Calculate the parameters of the decision function $y(x) = x^\top Ax + b^\top x + c$ of the 'quadratic discriminant analysis' that can be derived by doing classification in a Gaussian model *without* assuming that $\Sigma_+ = \Sigma_-$, and calculate the training- and the test-error rate of this algorithm.
   (d) [optional] For each data-point in the test-set, calculate its (scaled and signed) distance to the decision boundary (i.e. the values of $y(x)$ for each x). Make a plot which contains the histogram of all points in the positive class (in blue) as well as a histogram of the points in the negative class (in red).
   (e) [optional] Calculate the decision boundary of the quadratic algorithm and add it to the plot used in (b).