Solutions to exercises: week 1

Exercise 1.1
(a) See book, pg 19: $x_0 = 1/2$.
(b) Not really well-defined: in principle the decision boundary is everywhere where $p(\omega_1|\mathbf{x}) = p(\omega_2|\mathbf{x})$, but in many locations this results in 0/0, so no decision boundary is defined. In practice all x between 1 and 2 can be used to switch your decision from ω_1 to ω_2.
(c) Strictly speaking all points between 0.5 and 1 have equal posterior, and are therefore on the boundary.
(d) At $x_0 = 1$.

Exercise 1.2
(a) See book, pg 19: $x_0 = (1 - \log(2)/2$.
(b) The decision boundary is now all x between 0.5 and 1.

Exercise 1.3
(a) For $x = 3$, $p(\omega_1|x) = 0$ again, so assigned to class ω_2, for $x = -0.5$, $p(\omega_1|x) = 1$, so assigned to class ω_1. For $x = 0.5$ it is a bit more difficult. Note that for the overlap area between the classes holds that $p(x|\omega_1) = 1 - x$ and $p(x|\omega_2) = x/4$. So for $x = 0.5$ we have $p(x|\omega_1) = 0.5$ and $p(x|\omega_2) = 0.5 * 0.25 = 0.125$. Therefore $p(\omega_1|x) = \frac{1/2}{1/2+1/8} = 4/5$.
(b) The output labels flip classes at $x = 4/5$, so there is the decision boundary.

Exercise 1.4
(a) Only $x = 0.5$ changes, and it becomes $p(\omega_1|x) = 2/3$.
(b) At $x = 2/3$.

Exercise 1.5
Just take the overlap area where objects of class ω_1 and ω_2 go wrong: $\varepsilon^* = P(\omega_2) * 1/2 * (4/5 * 1/5) + p(\omega_1) * 1/2 * (1/5 * 1/5) = 1/20$.

Exercise 1.6
(b) The `mean(x)` gives the average over all elements, `mean(x,axis=0)` the average over objects (thus the average object), while `mean(x,axis=1)` gives the average value of the features of one object (thus pretty useless).

Exercise 1.7
(b) This really depends on your data and your imagination.

Exercise 1.11
(a) 0, 0, 0, negative.
(b) Positive.

Exercise 1.12
(a) An ellipsoid that is aligned with the axes of the plot with the long axis of the ellipsoid, which is along the y-axis, 3 times larger than the short ellipsoid axis. Ratio is 3.
(b) Covariances increase. Largest variance decreases, smallest goes up.
(c) Drawing should be straight diagonal line, e.g. /. If the variances are a and b, the covariance should be $\sqrt{ab}$.

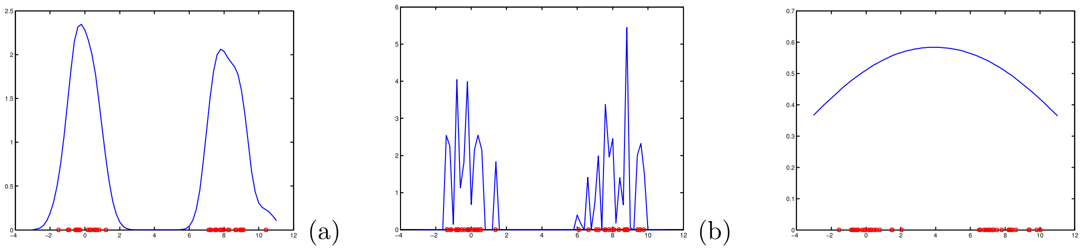
Exercise 1.14
(a) Only linear classifiers are possible.

Exercise 1.15
(a) Linear.
(b) In a way that doesn't depend on the optimal boundary. Typically the shape will be quadratic because of the limited training points.
(c) ldc is by construction linear, qdc will typically remain nonlinear due to the finite sample size.

(d) The solutions by ldc and qdc will coincide.

Exercise 1.17

(d) The data consists of two Gaussian clusters with $\sigma = 1$. The width parameter should therefore be around 0.5 to 1. In figure (a) the estimate for $h = 0.5$ is shown while figures (b) and (c) depict the estimates for $h = 0.05$ (too small) and $h = 5$ (too large), respectively.



(a) (b) (c)

Exercise 1.18

(a) The log-likelihood increases as h decreases. In the plot in figure (a) below, h varies from 0.05 to 5, and the highest log-likelihood is achieved at $h = 0.05$. This is due to the fact that the log-likelihood is computed on the training set.

Exercise 1.19

(a) These curves are depicted in figure (b) below, where the test set curve has a clear maximum at $h = 1$ (and not at the minimal h as in the training set curve. In the training set curve there is a clear over-training effect.

Exercise 1.20

(c) The quadratic, linear and Fisher discriminant are not affected, the nearest mean and the k-NN are. It is because in the first classifiers a covariance matrix $\hat{\Sigma}$ is estimated. Therefore automatically the scale of the feature values is estimated in these classifiers.

(d) In many cases it is an advantage, because it means that you don't need to optimize the scaling of the features. The disadvantage is that the training set should be large enough to make the estimation of the scale reliable.