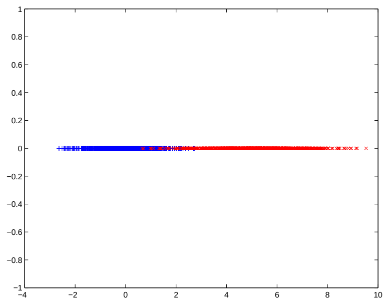
Solutions to exercises of Day 1

Exercise 1.10

The `gendatk` and `gendatp` routines assume local distributions of the data. Based on the small number of samples given, they cannot generalise the underlying distribution well.

Exercise 1.11



Exercise 1.12

(a) 0.5.
(b) A Gaussian probability density function.
(c) $\mu = 0.5$ and $\sigma^2 = \frac{1}{12n} \approx 0.0833/n$ (remember the definition of the standard error of the mean).
(d) The mean will be approximately the same, the variance will be lower.
(e) Nothing changes, only the histogram may become slightly more smooth. We have just sampled a bit more from the Gaussian; the individual estimates haven't improved.

Exercise 1.14

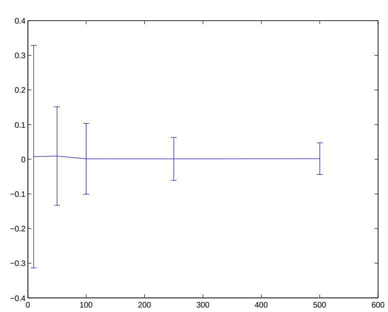
They are random. Because the dataset is spherical, any set of eigenvectors will do.

Exercise 1.15

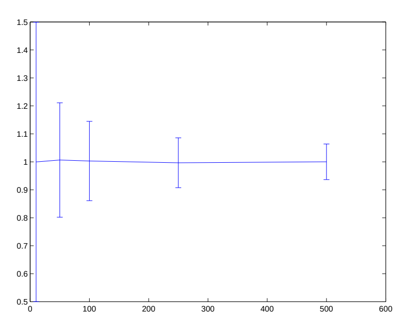
(a) Now you should find unit eigenvectors aligned with the axes, and eigenvalues approximately 3 and 1.

Exercise 1.17

(a) The resulting graphs should look like this (mean in figure (a), standard deviation in (b):)



(a)



(b)

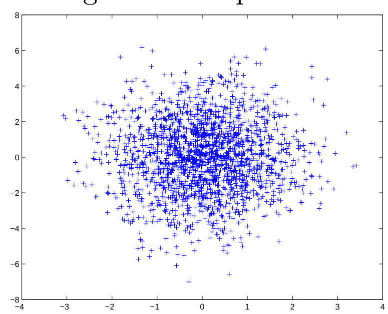
(b) For these settings, nothing much seems to improve for more than 250 samples.

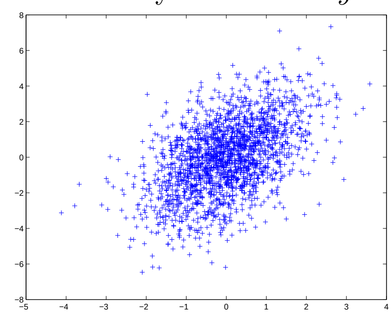
Exercise 1.18

(a) The condition number should increase rapidly for more than 7 or 8 dimensions.

Exercise 1.19

(b) The diagonal elements are the variances of each of the variables (in this case there are two variables, say, x and y). The off-diagonal elements represent the covariances. For the first case, `Sigma = [1 0; 0 4]`, the covariance of x and y is zero. In figure (a), this results in a Gaussian distribution which is aligned with the axes. For the second case, `Sigma = [1 1; 1 4]`, x and y do co-vary, which is visible in the fact that the distribution is oriented at an angle with respect to the axes. We can also say that x and y are correlated.

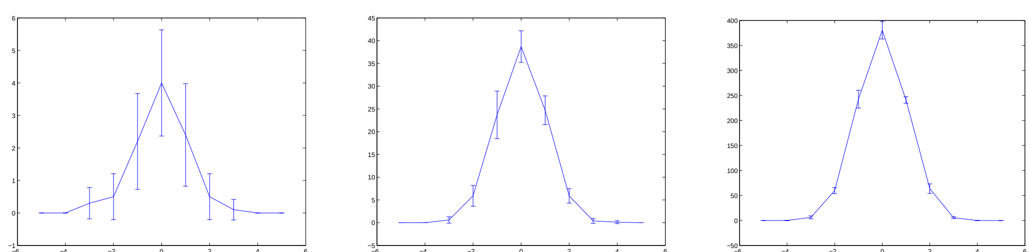

(a)


(b)

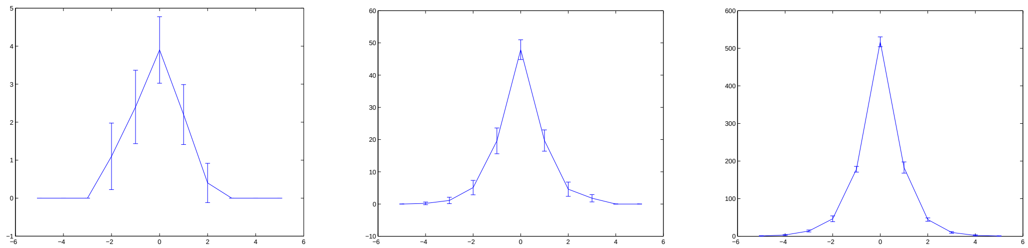
(c) The matrix is not symmetric: covariance of x and y is exactly the same as the covariance of y and x.
(d) The variance of the second variable, y, is much larger than the variance of x. Consequently, the distribution is stretched along the axis corresponding to y.

Exercise 1.20

(a) The output should look like this:



(b) With the script given, at least 1000 samples are needed.
(c) The same holds in general, but the difference is that for the Laplacian the mass is more concentrated around the mean and can therefore be estimated more easily (the histogram's standard deviation is lower w.r.t. its mean than for the Gaussian):



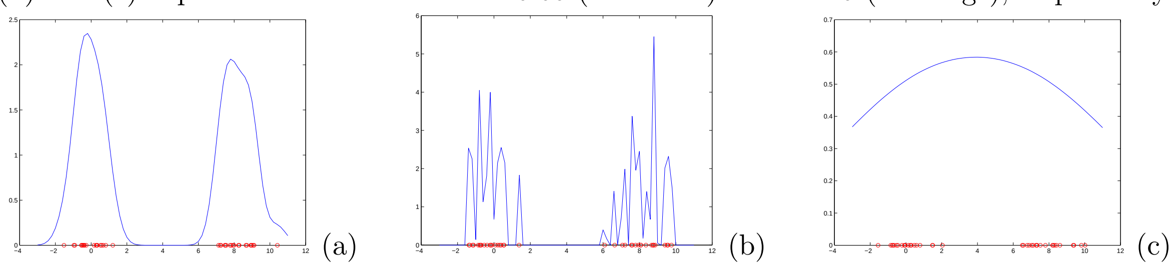
The errors in the tails are significantly larger, though.

Exercise 1.21

(a) Judged visually, $\mathcal{O}(10^4)$.
(b) For smaller bin sizes the variation between different realisations of the histograms is much larger. To get a stable result, for 5×5 bins, $\mathcal{O}(10^3)$ are needed; for 20×20, $\mathcal{O}(10^4 - 10^5)$.

Exercise 1.22

(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.



Exercise 1.23

(a) The log-likelihood increases as h decreases. In the plot in figure (a) below, h varies from 0.01 to 5, and the highest log-likelihood is achieved at $h = 0.01$. This is due to the fact that the log-likelihood is computed on the training set.

Exercise 1.24

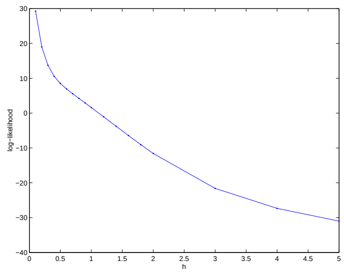
(a) These curves are depicted in figure (b) below, where the test set curve has a clear plateau between $h = 0.5$ and $h = 1$ (and not at $h = 0.01$ as in the training set curve). In the training set curve there is a clear over-training effect.

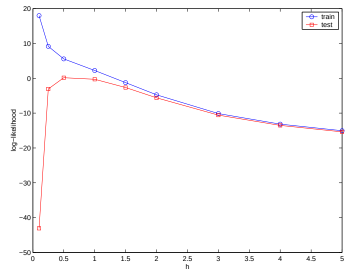
Exercise 1.25

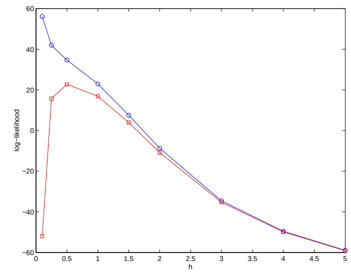
(a) It should be roughly in the same order.

Exercise 1.26

For more training examples, there is a clearer preference for a smaller width parameter. This is clearly visible in figure (c) below.



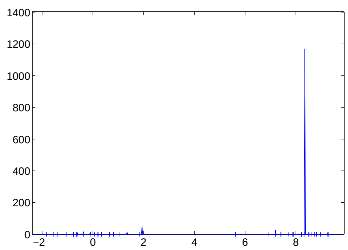
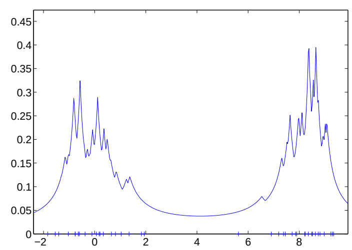
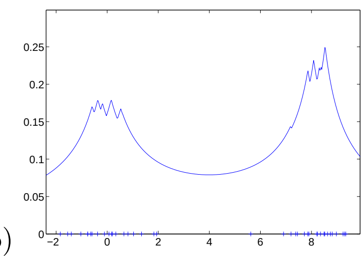
(a)


(b)


(c)

Exercise 1.27

(a) “Infinite peaks” around each sample; as the distance to each training sample goes to zero, the estimated density goes to infinity. You don’t see that because we only calculate the density at certain grid points.
(c) Some example are plotted below: (a) $k = 1$, (b) $k = 5$ and (c) $k = 15$. For $k = 5$, the density looks reasonable.


(a)

(b)

(c)

(d) Using cross-validation, just as for h in Parzen density estimation.