MLPR: Lab 3

Nicholas Piël
cknr: 0104612

2007-09-27

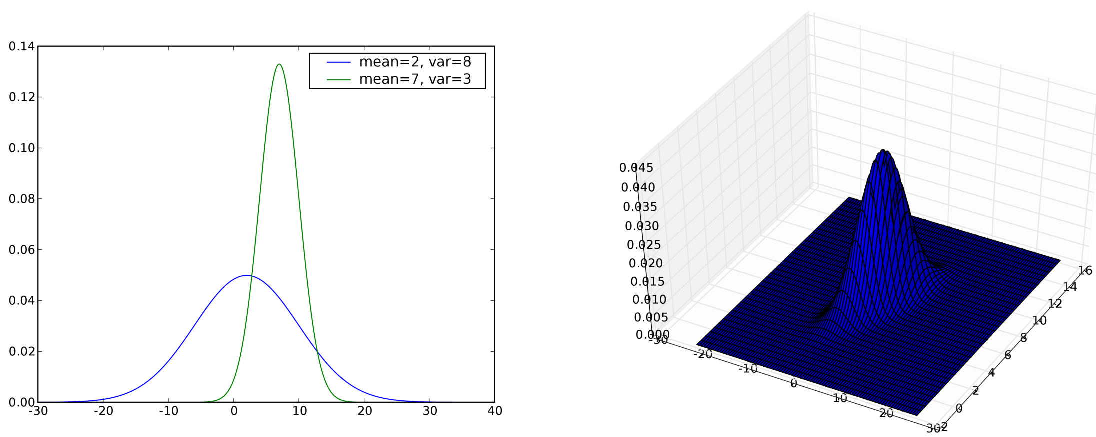
Abstract

This lab is focused on the parameter estimation of a normal distribution. In Machine Learning it is very important to visualize your data and your distri- butions in order to get a better insight of the algorithms. This lab has two sections. In the first, we create a visualization tool. In the second we create a program able to learn the parameters of a given dataset and we start experimenting with different priors. abstract

1 Visualisation Tool

You can find the visualisation of the distribution of the values in the table below in figure 1. The code which generated this plot works can be found in listing 1.

Distribution	1^{st}	2^{nd}	3^{nd}
mean	2	7	2,7
σ	8	3	8,3;3,3



(a) 2D Visualisation of two different distributions

(b) a 3D visualisation of one multidimensional distribution

Figure 1: Visualisation of various distributions

1.1 Discussion

One can see that for visualising the distribution of a 1D dataset we already need a 2D picture. A 2D dataset already needs a 3D visualisation. While we could even add another dimension by adding color, Listing 1: Visualisation code

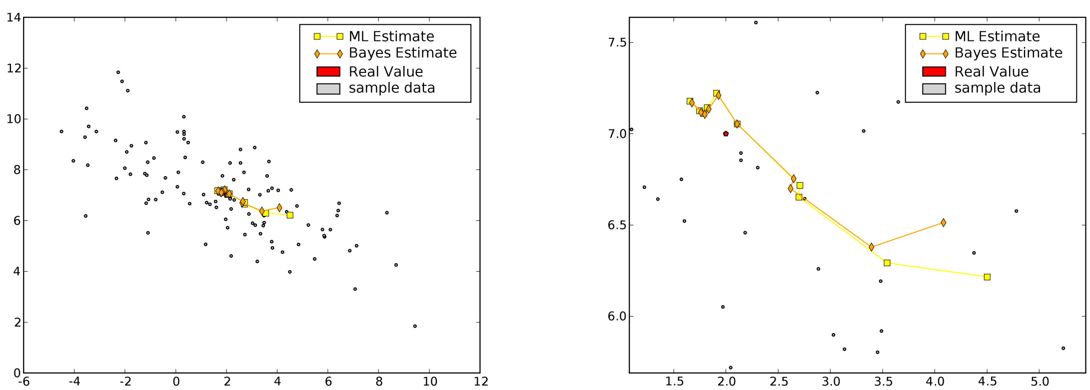
```
function plotSurf(mu, sigma):
    if len(mean) == 2: # 2D Plot
        minx = mu[0] - 3*sigma[0][0]
        miny = mu[1] - 3*sigma[1][1]
        maxx = mu[0] + 3*sigma[0][0]
        maxy = mu[1] + 3*sigma[1][1]
        X = range(minx, maxx, 0.05)
        Y = range(miny, maxy, 0.05)
        Z = gaussDistribution(mu, sigma, X, Y)
        plot_surface(X,Y,Z)
    elif len(mean) == 1: # 1D Plot
        span = sigma1 * 3
        from = mean - span
        to = mean + span
        X = arange(from,to,0.1)
        Y = gauss(mu, sigma, r)
        plot(X,Y)
```

this leaves us wondering how to appropriately visualise even more dimensions.

From plot 1A we can see that a lower variance will result in a lower peak. The mean selects around which point the distribution is built. The plots also strengthen our belief that the area under surface should sum up to one.

2 Parameter Estimation

Now its time to experiment, we pick the same mean and variance as in the above situation and sample 100 points from this distribution. The scatterplot of this distribution can be found in figure 2. We have also plotted the real mean within this scatterplot as can be found in figure 2. The distribution of this dataset can be found in figure 1b.



(a) The Gaussian distribution

(b) Zoomed in segment of the distribution

Figure 2: Scatter plot of 100 points sampled from a Gaussian distribution, priors of the Bayesian distribution set to the original values from which we created the distribution: mean = [2,7], var = [[8,3],[3,4]]

The plots show the distribution and the estimation of the mean by using the ML estimate method and the Bayesian approach. I have connected the estimated means of the two approaches with a line in different color. From this we can already see that both methods approach the real mean when more data arrives. The difference between both approaches is best visible when we have a small set of data over which we try to estimate. For example when we only estimate the mean on 10 datapoints the Bayesian approach does far better at approaching the real mean then the ML Estimate approach, however this difference becomes negligible after 40 datapoints.

Now we start playing with the prior so that we can see and understand what happens with the estimated mean over the sampled data. We start by changing the covariance from very low (meaning that we are pretty sure over our initial estimate)to high (meaning that we are uncertain about our priors). We leave the prior for the mean at [2,7]. The result can bee seen in figure 3

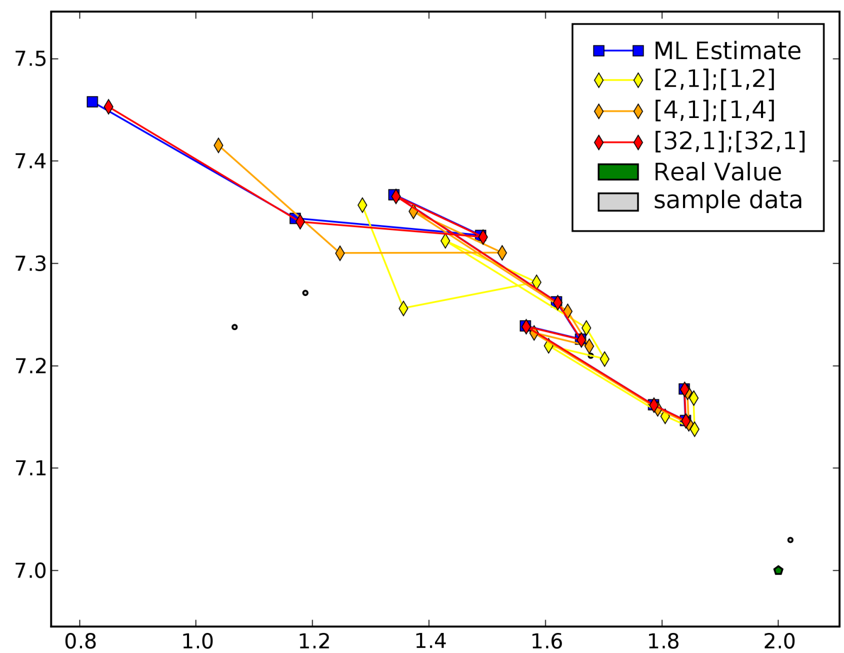


Figure 3: Estimating the mean over Gaussian data with different covariances

In figure 3 we can again see that all methods approach the real mean (the green dot at the bottom right) when more data arrives. When we set the covariance matrix to something high we see that the Bayesian estimator behaves almost the same as the ML estimater, we can interpret this in such a way that a Bayesian Estimater with very low beliefs in its priors equals the ML approach. The higher the belief of its priors (the smaller its covariance) the more it weights in its priors and, in this case the more correctly it estimates the mean.

Since in this situation the mean was set to the 'real value' it will be interesting to see what happens when we play with those values. So for our next experiment we set covariance to something small indicating a high belief in its prior and vary over the prior mean. The result can be found in figure 4

From the plot in figure 4 we can clearly see a different approach for all 4 estimation methods. We can see that the red and orang line which had a belief which was very offset (-10,-10) and (0,0) have their beginning estimates at a completely different location then the ML estimate or the Bayesian estimator with a 'correct' prior.

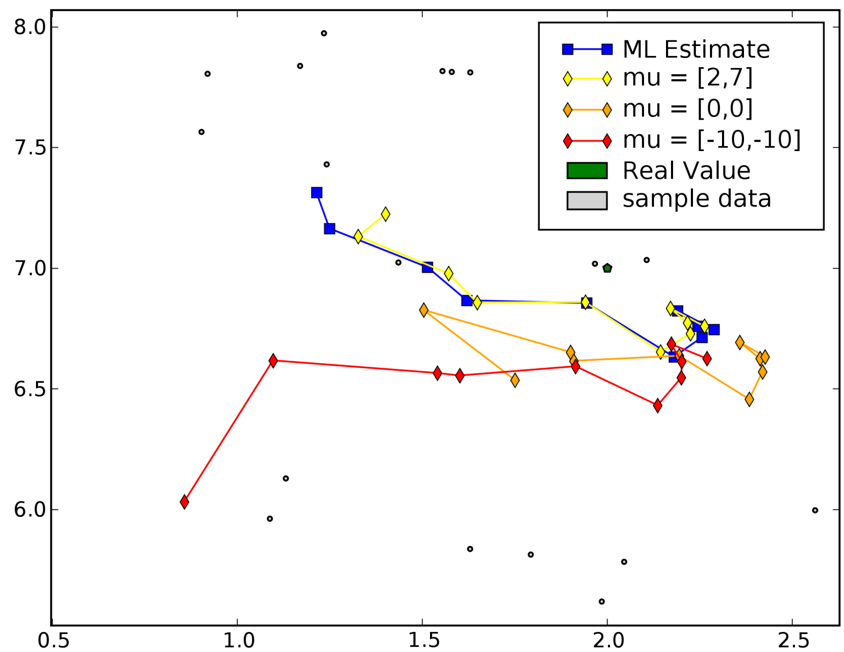


Figure 4: Scatter plot of 100 points sampled from a Gaussian distribution, priors of the Bayesian distribution set to [[2,1];[1,2]] for the variance and it varies for the mean.