

# Stat 406 Spring 2008 Homework 1

## 1 Normalization constant for a 1D Gaussian

The normalization constant for a zero-mean Gaussian is given by

$$Z = \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \quad (1)$$

where  $a = -\infty$  and  $b = \infty$ . To compute this, consider its square

$$Z^2 = \int_a^b \int_a^b \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) dx dy \quad (2)$$

Let us change variables from cartesian  $(x, y)$  to polar  $(r, \theta)$  using  $x = r \cos \theta$  and  $y = r \sin \theta$  (see handout on change of variables). Since  $dx dy = r dr d\theta$ , and  $\cos^2 \theta + \sin^2 \theta = 1$ , we have

$$Z^2 = \int_0^{2\pi} \int_0^\infty r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr d\theta \quad (3)$$

Evaluate this integral and hence show  $Z = \sigma\sqrt{2\pi}$ . Hint 1: separate the integral into a product of two terms, the first of which (involving  $d\theta$ ) is constant, so is easy. Hint 2: if  $u = e^{-r^2/2\sigma^2}$  then  $du/dr = -\frac{1}{\sigma^2} r e^{-r^2/2\sigma^2}$ , so the second integral is also easy (since  $\int u'(r) dr = u(r)$ ).

## 2 Normalization constant for a multidimensional Gaussian

Prove that the normalization constant for a  $d$ -dimensional Gaussian is given by

$$Z = (2\pi)^{d/2} |\Sigma|^{-\frac{1}{2}} \quad (4)$$

In other words, show that

$$\int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \Sigma) d\mathbf{x} = 1 \quad (5)$$

You may assume  $\boldsymbol{\mu} = \mathbf{0}$  for simplicity. Hint: diagonalize  $\Sigma$  and use the fact that  $|\Sigma| = \prod_i \lambda_i$  to write joint pdf as a product of  $d$  one-dimensional Gaussians in a transformed coordinate system. (You will need the change of variables formula.) Finally, appeal to the result for univariate Gaussians.

## 3 2d Gaussians: plotting and whitening (Matlab)

1. Load the data in `heightWeightDataSimple.txt` using `dload`. The first column represents sex (1=male, 2=female), column 2 is height (in feet), column 3 is weight (in pounds).
2. Make a scatter plot of the data. Use red circles for the females, blue crosses for the males. Turn in your figure and code.
3. Fit a 2d Gaussian to each group of data, using the empirical mean and covariance:

```
mu = mean(X(ndx, :));  
S = cov(X(ndx, :));
```

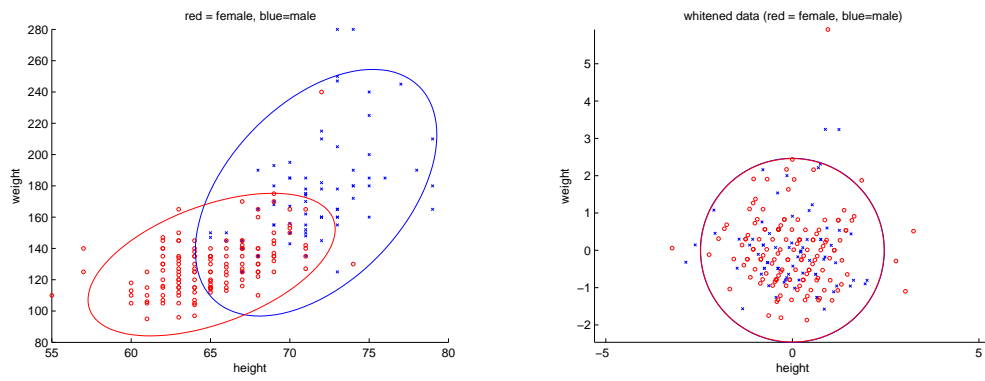


Figure 1: Height-weight data. Left: original. Right: whitened.

where `ndx` selects the rows corresponding to men or women. Plot your Gaussians as ellipses (use `gaussPlot2d`), superimposing on your scatter plots. It should look like Figure 1(left). Turn in your figure and code.

4. Now whiten each group of data separately. Compute the empirical mean and covariance on the new data, and plot the whitened data and the new Gaussians. It should look like Figure 1(right). Turn in your figure and code.