LFD 2005 Tutorial Solutions - Week 4

October 16, 2006

1 Change of variables of Guassian Distribution

By plugging in the change $y = (x - \mu)/\sigma$ we get a function of the form of a zero mean, unit variance Gaussian. However the constant factor is not correct. That is because we are making a change of variables in a distribution - and a distribution must integrate to one. In changing variables, you are changing the associated lengths/hypervolumes in the space and so you need to adjust for the changes. For example a uniform distribution between 0 and 1 becomes a uniform distribution between 0 and 10 using $y = 10x$. But each region $(x, x + \delta x)$ is also ten times as long, and so we need to reduce the height by a factor of 10 to keep the area the same.

Doing this for our Gaussian distribution, we find we need to multiply all the values by σ giving us the required distribution.

Generally:

$$\int_S f(\mathbf{x})\ d\mathbf{x} = \int_S f(\mathbf{y}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|\ d\mathbf{y} \tag{1}$$

where we have used the shorthand

$$\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right)_{ij} = \frac{\partial x_i}{\partial y_j} \tag{2}$$

and $|.|$ denotes the determinant (det() in matlab). The quantity

$$\left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| \tag{3}$$

is called the Jacobian, and basically gives the ratio of hypervolumes in the first coordinate system to hypervolumes in the other.

Using this, for probability density $P(\mathbf{y})$ we have $P(\mathbf{y}) = P(\mathbf{x})|\partial\mathbf{x}/\partial\mathbf{y}|$.

For a one dimensional problem the Jacobian reduces to the absolute value of a derivative.

2 ML Gaussian Parameter Estimation

Likelihood of the data, $\mathbf{x}$, being generated from a Gaussian $N(x; u, \sigma^2)$, is

$$
\begin{aligned}
L(\mathbf{x}|\mu, \sigma^2) &= \prod_{i=1}^{P} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\
&= \frac{1}{(2\pi^2)^{\frac{P}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{P}(x_i-\mu)^2} \\
LL(\mathbf{x}|\mu, \sigma^2) &= -\frac{P}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{P}(x_i-\mu)^2
\end{aligned}
$$

Find the estimator for σ^2 which maximizes the log likelihood by differentiating and equating to zero

$$
\begin{aligned}
\frac{\partial LL}{\partial \hat{\sigma^2}} &= -\frac{P}{2\hat{\sigma^2}} - \frac{1}{2(\hat{\sigma^2})^2}\sum_{i=1}^{P}(x_i - \hat{\mu}) \\
\hat{\sigma^2} &= \frac{1}{P}\sum_{i=1}^{P}(x_i - \hat{\mu})
\end{aligned}
$$

3 Class Conditional Classification

Classification probability is given by Bayes theorem;

$$
\begin{aligned}
p(c_j|x) &= \frac{p(x|c_j)p(c_j)}{p(x)} \\
&= \frac{p(x|c_j)p(c_j)}{\sum_{k=1}^{N} p(x|c_k)p(c_k)}
\end{aligned}
$$

In this case, using the using the Gaussian likelihood $p(x|c_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}(\frac{x-\mu_j}{\sigma_j})^2}$ and ML parameter estimators $\hat{\mu_j} = \frac{1}{N_j}\sum_{i=1}^{N_j} x_i$, $\hat{\sigma_j^2} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_i - \hat{\mu_j})$, $\hat{p}(c_j) = \frac{N_j}{\sum_{k=1}^{N} N_k}$, we have; $\hat{\mu_1} = 0.26$, $\hat{\mu_2} = 0.8625$, $\hat{\sigma_1^2} = 0.0149$, $\hat{\sigma_2^2} = 0.0092$, $\hat{p}(c_1) = 0.714$, $\hat{p}(c_2) = 0.2857$ and $p(c_1|0.6) = 0.6305$.

4 Matlab Exercise

See attached w4.m

5 Decision Boundary

Decision boundary is given by the value(s) of x where the posterior probability of each class is equal

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2} p_1 \quad = \quad \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}(\frac{x-\mu_2}{\sigma_2})^2} p_2$$

Taking logs and rearranging

$$-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2 - -\frac{1}{2}(\frac{x-\mu_2}{\sigma_2})^2 + ln(\frac{p_2\sigma_1}{p_1\sigma_2}) = 0$$

When expanded gives a quadratic in x where

$$\begin{aligned} ax^2 + bx + c &= 0 \\ a &= \sigma_1^2 - \sigma_2^2 \\ b &= 2(\sigma_2^2\mu_1 - \sigma_1^2\mu_2) \\ c &= \sigma_2^2\mu_1^2 + \sigma_1^2\mu_2^2 - 2\sigma_1^2\sigma_2^2 ln(\frac{p_2\sigma_1}{p_1\sigma_2}) \end{aligned}$$

Solutions are given by the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Hence there can be one, two or zero decision boundaries, (consider the cases where $\mu_1$and μ_2 are are far, where $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 > \sigma_2^2$, and the latter but where in addition, $p_2/p_1 < \sigma_2^2/\sigma_1^2$).