Assignment 1

Deep Learning 1, SS22

Team Members		
Last name	First name	Matriculation Number
Amering	Richard	1331945

1 Task 1 - Maximum Likelihood Estimation

1.1 Derivation

Likelihood of a single sample of dimension m:

$$p(x|\theta) = \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

Likelihood for a set of N samples:

$$X = \{x_1, x_2..x_N\}$$

$$p(X|\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}$$

$$|\Sigma| = \sigma^6$$

Add logarithms to both sides of the equation to obtain log-likelihood:

$$log(p(X|\theta)) = \sum_{i=1}^{N} log(\frac{1}{\sqrt{(2\pi)^3|\Sigma|}} e^{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)})$$

$$log(p(X|\theta)) = \sum_{i=1}^{N} log(\frac{1}{\sqrt{(2\pi)^3|\Sigma|}}) + log(e^{-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)})$$

$$log(p(X|\theta)) = \sum_{i=1}^{N} log(\frac{1}{\sqrt{(2\pi)^3|\Sigma|}}) - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = (x_i-\mu)^T\sigma^{-2}I^{-1}(x_i-\mu) = \sigma^{-2}(x_i-\mu)^T(x_i-\mu)$$

Inner product given by:

$$(x_i-\mu)^T(x_i-\mu) = \sum_{j=1}^{m}(x_{i,j}-\mu_{i,j})^2$$

1.2 Finding the Maximum

For finding the maximum likelihood estimation of μ, the log-likelihood is derived wrt. the different vector components of μ, and then set to zero. Since the log-expression is monotonically increasing and without upper bound, a root in the derivative of the expression must be corresponding to a minimum in the log function, and therefore a minimum in the original likelihood function. This is because a monotonic function without upper bounds like the likelihood function does not have maxima.

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{d}{d\mu_j}(x_{i,j} - \mu_j)^2$$

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{d}{d\mu_j}(x_{i,j}^2 - 2x_{i,j}\mu_j + \mu_j^2)$$

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{\sigma^2}\sum_{i=1}^{N}(-x_{i,j} + \mu_j)$$

$$\frac{dlog(p(X|\theta))}{d\mu_j} = -\frac{1}{\sigma^2}\sum_{i=1}^{N}(-x_{i,j}) + N\mu_j \overset{!}{=} 0$$

Finally giving the maximum likelihood estimation for μ, which is just the mean value of component j of samples x_i in the set:

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N}(x_{i,j})$$