Machine Learning Sheet 2

1) $p(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$. Give maximum likelihood for N datapoints $x_1, .., x_N$.

Maximize Likelihood $L(\lambda) = \prod_{n=1}^{N} p(x^{(n)} | \lambda)$ with respect to the parameter λ for the given datapoints.

Since $L(\lambda) \geq 0$ and $\log(y)$ $\overset{y \in \mathbb{R}}{\text{is}}$ strictly monotonously growing, it is true that: $L(\lambda)$ maximized $\Leftrightarrow$ $\log(L(\lambda))$ maximized.

Maximize: $\mathcal{L}(\lambda) = \log(L(\lambda))$

$$\mathcal{L}(\lambda) = \sum_{n=1}^{N} \left(x^{(n)} \log(\lambda) - \log(x^{(n)}!) - \lambda \right)$$

$$\frac{\partial}{\partial \lambda} \mathcal{L} \overset{!}{=} 0 \quad \Rightarrow \quad \sum_{n=1}^{N} \left(\frac{x^{(n)}}{\lambda} - 1 \right) = 0$$

$$\Leftrightarrow \quad \sum_{n=1}^{N} \left(x^{(n)} - \lambda \right) = 0 \Leftrightarrow N\lambda = \sum_{n=1}^{N} x^{(n)}$$

$\Rightarrow$ The best estimate for λ is the mean of $x_1, .., x_n$.

2)

A: Every probability density function must be

a) Normalized. b) $p(x) \geq 0 \quad \forall x$

Here:

a) $$1 \overset{!}{=} \int_{-\infty}^{\infty} p(x)\,dx = \frac{1}{2} \int_{-\infty}^{\infty} \mathcal{N}(x | \mu_1, \sigma_1^2)\,dx + \frac{1}{2} \int_{-\infty}^{\infty} \mathcal{N}(x | \mu_2, \sigma_2^2)\,dx$$

$$= \frac{1}{2} + \frac{1}{2}$$

b) $\mathcal{N} \geq 0 \Rightarrow p \geq 0$

B: For the derivation of the maximum likelihood, maximum of L the consideration of $\log(L)$ instead of L has proven to be useful and easier than L directly. However, in this special case, $\log\left(\frac{1}{2}\mathcal{N}_1 + \frac{1}{2}\mathcal{N}_2\right)$ as the logarithm of a sum is tedious to compute, especially the derivative will produce many terms.