University of Waterloo
CS 480
r2knowle: 2023-09-20

Exercise # 5

Q1) We are going to start with the following Poisson distribution:

$$Pr(Y_i = k|X_i) = \frac{\mu_i^k}{k!}\exp(-\mu_i)$$

We can then take the log of this to get:

$$\begin{aligned} L(\mu_i) &= \prod_{i=0}^{n} \frac{\mu_i^k}{k!}\exp(-\mu_i) \\ l(\mu_i) &= \log(\prod_{i=0}^{n} \frac{\mu_i^k}{k!}\exp(-\mu_i)) \\ &= \sum_{i=0}^{n} \log(\frac{\mu_i^k}{k!}\exp(-\mu_i)) \\ &= \sum_{i=0}^{n} \log(\mu_i^k) - \log(k!) - \mu_i \\ &= \sum_{i=0}^{n} k\log(\mu_i) - \log(k!) - \mu_i \end{aligned}$$

Q2) Parameter pi represents the mean and also the probability of a given event in the Bernoulli distribution. Therefore pi needs to be a number between 0 and 1. In order to accomplish this we do a logit transform to get:

$$\log \frac{p_i}{1-p_i} = w^T x_i + b$$

However for μ_i we can have any possible value as its only the mean and not the probability, because the actual probability is bounded by the function given in (6) we can do a logit transform where we just take in the value. Thus we can the transform of:

$$\log \mu_i = w^T x_i + b$$

Q3) We would like to optimize and maximize the value of μ, thus we can transform log likelihood:

$$\begin{aligned} max_\mu\left(l(\mu_i)\right) &= max_\mu\left(\sum_{i=0}^{n} k\log(\mu_i) - \log(k!) - \mu_i\right) \\ &= max_\mu\left(\sum_{i=0}^{n} k(w^T x_i + b) - log(k!) - e^{w^T x_i + b}\right) \end{aligned}$$

Since k is a constant, we we now only have variables in terms of b and w we get:

$$max_{w,b}\left(l(w,b)\right) = max_\mu\left(\sum_{i=0}^{n} k(w^T x_i + b) - e^{w^T x_i + b}\right)$$

Q4) To calculate the weight vector and b we will first calculate the gradients. To start we will take the gradient $w.r.t$ to w:

$$\begin{aligned}
\frac{\delta}{\delta w} &= \sum_{i=0}^{n}(k(w^T x_i + b) - e^{w^T x_i + b})' \\
&= \sum_{i=0}^{n}(kx_i - x_i e^{w^T x_i} e^b) \\
&= \sum_{i=0}^{n} kx_i - e^b \sum_{i=0}^{n} x_i e^{w^T x_i} \\
&= k\sum_{i=0}^{n} x_i - e^b \sum_{i=0}^{n} x_i e^{w^T x_i}
\end{aligned}$$

This gives us our value of the gradient we will then introduce a learning factor α and update w by:

$$w \mathrel{+}= \alpha \frac{\delta}{\delta w}$$

Moving on to b, we will take the gradient $w.r.t$ to b:

$$\begin{aligned}
\frac{\delta}{\delta b} &= \sum_{i=0}^{n}(k(w^T x_i + b) - e^{w^T x_i + b})' \\
&= \sum_{i=0}^{n}(k - e^{w^T x_i} e^b) \\
&= \sum_{i=0}^{n} k - e^b \sum_{i=0}^{n} e^{w^T x_i} \\
&= kn - e^b \sum_{i=0}^{n} e^{w^T x_i}
\end{aligned}$$

This gives us our value of the gradient we will then introduce a learning factor α and update b by:

$$b \mathrel{+}= \alpha \frac{\delta}{\delta b}$$