Assignment 1 - Q3b
CS698 - Winter 2017

Vineet John (v2john@uwaterloo.ca)

1 Problem Statement

Show that this objective is equivalent to the negative log-likelihood for linear regression where each data point may have a different Gaussian measurement noise. What is the variance of each measurement noise in this model?

$$L(w, b) = \sum_{n=1}^{m} r_n(y_n w x_n + b)^2$$

2 Solution

Assume that linear regression problem is modeled such that $w^T \bar{x}_n - b$ is the prediction term rather than just $w^T \bar{x}_n$ as in the above equation.

Given that $y = f(x) + \epsilon$, where ϵ is the measurement error and $f(x)$ is the underlying function to be predicted, $f(x) = w^T \bar{x}_n - b$.
The measurement noise ϵ is assumed to be a Gaussian distribution such that $\epsilon = N(0, \sigma^2)$

Now, the probability of the target variable y, given the input parameters X, w and b can be denoted as:

$$Pr(y|\bar{X}, w, b, \sigma) = N(y|w^T \bar{x}_n - b, \sigma)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y - w^T \bar{x}_n + b)^2}{2\sigma^2}}$$

To obtain the optimal value of w and b, the maximum likelihood estimate is evaluated:

$$w^*, b^* = argmax_{w,b} Pr(y|\bar{X}, w, b, \sigma)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(y - w^T \bar{x}_n + b)^2}{2\sigma_n^2}}$$

Here, σ_n denotes that each of the data-points has a different Gaussian measurement associated with it.

Evaluating, the log of the maximum likelihood equation above,

$$w^*, b^* = argmax_{w,b} \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_n} * (-\frac{(y - w^T \bar{x}_n + b)^2}{2\sigma_n^2})$$

Now, discarding the constant terms $\frac{1}{\sqrt{2\pi}}$ and $\frac{1}{2}$,

$$w^*, b^* = argmax_{w,b} \sum_{n=1}^{N} \frac{1}{\sigma_n} * (-\frac{(y - w^T \bar{x}_n + b)^2}{2\sigma_n^2})$$

$$= argmax_{w,b} \sum_{n=1}^{N} \frac{1}{\sigma_n} * (-\frac{(y - w^T \bar{x}_n + b)^2}{\sigma_n^2})$$

$$= argmax_{w,b} \sum_{n=1}^{N} -\frac{(y - w^T \bar{x}_n + b)^2}{\sigma_n^3}$$

Here, the standard deviation term σ_n^3 could be said to represent r_n, which models a distinct weight for each of the data-points.

Consequently, the final expression could be written as:

$$w^*, b^* = argmin_{w,b} \sum_{n=1}^{N} r_n (y - w^T \bar{x}_n + b)^2 \qquad (1)$$

which is the originial linear regression loss function, obtained from the negative log-likelihood of linear regression where each point has a different Gaussian measurement, with $\mu = 0$ and $variance = \sigma_i^2$ for a datapoint i.

The variance for this model can be evaluated by equating the assumption made in Equation 1.

$$\frac{1}{\sigma_n^3} = r_n$$

$$\sigma_n^3 = r_n^{-1}$$

Equation 2 gives the variance of each point in the model of linear regression where each point has a different Gaussian measurement.

$$\sigma_n^2 = r_n^{-2/3} \qquad (2)$$

3   Conclusion

The negative log-likelihood for linear regression, with each data point having a different Gaussian noise is proven to be the same as the objective with a different weight for each data point in the loss function.

Also, the variance of each point in the model is expressed in terms of the weight assigned in the weighted linear regression loss function.