Homework 3

Atlas42

September 2021

1 $X^T X$ is invertible when X is full rank

For X to be full rank, either when its rows or its columns are linearly independent. For example: If our matrix is an m $\times$ n matrix with $m < n$, then it has full rank when its m rows are linearly independent or Rank $(X) = m$. If $m > n$, the matrix has full rank when its n columns are linearly independent or Rank $(X) =$ n. If $m = n$, the matrix has full rank either when its rows or its columns are linearly independent (when the rows are linearly independent, so are its columns in this case).

Assume that $m \leq n$ and Rank $(X) =$ m, and let $X^T X_u = 0$ for some $u \in R^m$ We need to show that $u = 0$, We also have that

$$0 = (X^T X_u, u) = (Xu, Xu),$$

and thus $X_u = 0$. But as Rank $(X) = m$. (Otherwise, the columns of X would be linearly dependent, and hence its rank less than m.) Assume that $X^T X \in R^{m\times m}$ is invertible. Then $m =$ Rank $(X^T X) \leq$ Rank $(X) \leq \min\{m, n\}$. Thus $\min\{m, n\} = m$, Rank $(X) = m$ and $m \leq n$

2 Proof $t = y(x, w) + noise- > w = (X^T X) - 1X^T t$

Suppose that the observations are drawn independantly from a Gaussian distribution:

$$t = y(x, \mathbf{w}) + \mathcal{N}(0, \beta^-1)t = \mathcal{N}(y(x, \mathbf{w}), \beta^-1)$$

with Precision parameter: $\beta = \frac{1}{\sigma^2}$

We now use the training data x, t to determine the values of the unknown parameters w and by maximum-likelihood. If the data are assumed to be drawn independently from the distribution then the likelihood function:

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, w), \beta^-1)$$

It is convenient to maximize the logarithm of the likelihood function:

$$\log p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \sum_{n=1}^{N} (\log \mathcal{N}(t_n | y(x_n, w, \beta^-1))$$
$$= -\frac{\beta}{2} \sum_{n=1}^{n} (y(x_n, \mathbf{w} - t_n))^2 + \frac{N}{2} \log \beta - \frac{N}{2} \log 2\pi$$

max $\log p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta = -\ \max\frac{\beta}{2} \sum_{n=1}^{N} (y(x_n, \mathbf{w}) - t)^2$

$= \min\ \frac{1}{2} \sum_{n=1}^{N} (y(x_n, \mathbf{w}) - t)^2$

Now, to find $\mathbf{w}$ we need to minimize $(y(x_n, \mathbf{w}) - t)^2$

Suppose:

$$\mathrm{L} = \frac{1}{2} \sum_{n=1}^{N} (y(x_n, \mathbf{w}) - t)^2$$

$$\mathrm{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ . & . \\ 1 & x_n \end{bmatrix},\ \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix},\ \mathrm{y} = \begin{bmatrix} 1 & y_1 \\ 1 & y_2 \\ \cdot & \cdot \\ . & . \\ 1 & y_n \end{bmatrix} = \begin{bmatrix} w_0 + x_1 W_1 \\ w_0 + x_2 W_1 \\ \cdot \\ . \\ w_0 + x_n W_1 \end{bmatrix} = \mathrm{xw}$$

$$\frac{\partial L}{\partial w} = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_0} \end{bmatrix} = \begin{bmatrix} t - xw \\ x(t - xw) \end{bmatrix} = x^{\mathbf{T}}(t - xw) = 0$$

$$x^{\mathbf{T}} t = x^{\mathbf{T}} xw$$
$$\mathrm{w} = (x^{\mathbf{T}} x)^{-1} x^{\mathbf{T} t}$$