01:640:495 - Lecture 6

Pranav Tikkawar

February 19, 2025

1. Suppose $(x_i, y_i), i = 1, \ldots, n$ are n (fixed) data points in $\mathbb{R}^2$. Our goal is to find a formula for θ_0, θ_1 such that the line $y = \theta_0 + \theta_1 x$ best fits these n points in the sense that it minimizes this quantity below

$$\sum_{i=1}^{n}(y_i - (\theta_0 + \theta_1 x_i))^2$$

i.e., if we are thinking of $\theta_0 + \theta_1 x_i$ as the predicted y_i at input x_i, then we are taking how far it is from the actual y_i, squaring that and adding over all i. In all parts, make sure to state/identify any technical assumptions and any corner/edge cases.

   (a) This quantity above is related to a distance between a fixed point and a point moving in a linear subspace of a Euclidean space. Identify the fixed point, moving point, the linear subspace, its basis and dimension, and the Euclidean space.

   Solution: Fixed point: $\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ Moving point is $\begin{bmatrix} \theta_0 + \theta_1 x_1 \\ \theta_0 + \theta_1 x_2 \\ \vdots \\ \theta_0 + \theta_1 x_n \end{bmatrix}$ This is the vector of predicted y_i values.
   Linear subspace: all vectors of the form

$$\vec{y} = \theta_0 \vec{1} + \theta_1 \vec{x}$$

   Our basis is $\{\vec{1}, \vec{x}\}$ where $\vec{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and the dimension is 2.
   Euclidean space: $\mathbb{R}^n$ where n is the number of data points.

   (b) Reach the goal.

Solution: To minimize

$$\sum_{i=1}^{n}(y_i - (\theta_0 + \theta_1 x_i))^2$$

We can notice that this is the same as the inner product of

$$\langle \vec{y} - \hat{y}, \vec{y} - \hat{y} \rangle$$

where $\hat{y} = \mathbf{X}\Theta = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$ We can say that this distance is minimized when $\vec{y} - \hat{y}$ is orthogonal to the column space of $\mathbf{X}$, i.e., when $\hat{y}$ is the projection of $\vec{y}$ onto the column space of $\mathbf{X}$.

$$\begin{aligned} X^T(\vec{y} - \hat{y}) &= 0 \\ X^T\vec{y} - X^T\hat{y} &= 0 \\ X^T\vec{y} - X^T X\Theta &= 0 \\ X^T\vec{y} &= X^T X\Theta \\ \Theta &= (X^T X)^{-1} X^T \vec{y} \end{aligned}$$

So we can say that the optimal Θ is given by

$$\Theta = (X^T X)^{-1} X^T \vec{y}$$

2. Suppose errors are weighted according to the x value - i.e., suppose $w(x)$ is a positive valued function and we would like to minimize

$$\sum_{i=1}^{n} w(x_i)(y_i - (\theta_0 + \theta_1 x_i))^2.$$

Do you think the method above still works? Explain. If yes, make the necessary adjustments.

Solution: We can do the same thing as above, redfine our notion of distance to be weighted distance.

$$\langle \vec{y} - \hat{y}, \vec{y} - \hat{y} \rangle = \sum_{i=1}^{n} w(x_i)(y_i - (\theta_0 + \theta_1 x_i))^2 = \sum_{i=1}^{n} w(x_i)(y_i - \hat{y})(y_i - \hat{y})$$

Now we want to minimize this which follows the same logic of the previous problem. We can say that this distance is minimized when $\vec{y} - \hat{y}$ is orthogonal to the column space of $\mathbf{XW}$, i.e., when $\hat{y}$ is the projection of $\vec{y}$ onto the column space of $\mathbf{XW}$.

$$\begin{aligned}
X^T W(\vec{y} - \hat{y}) &= 0 \\
X^T W\vec{y} - X^T W\hat{y} &= 0 \\
X^T W\vec{y} - X^T W X\Theta &= 0 \\
X^T W\vec{y} &= X^T W X\Theta \\
\Theta &= (X^T W X)^{-1} X^T W\vec{y}
\end{aligned}$$

3. Suppose we change how we measure error to using power 4:

$$\sum_{i=1}^{n} (y_i - (\theta_0 + \theta_1 x_i))^4.$$

Do you think the method above still works? Explain. If so, make the necessary adjustments.

Solution: No, Linearity is not conserved when trying to recreate a notion of distance and inner product.