DSC 40A

Theoretical Foundations of Data Science I

Last Time

- **Goal**: Find prediction rule $H(x)$ for predicting salary given years of experience.

- Minimize **mean squared error**:

$$\frac{1}{n}\sum_{i=1}^{n}\left(H(x_i) - y_i\right)^2$$

- To avoid **overfitting**, use linear prediction rule:

$$H(x) = w_1 x + w_0$$

In This Video

Which linear prediction rule minimizes the mean squared error?

Recommended Reading

Course Notes: Chapter 2, Section 1

Minimizing the MSE

- The MSE is a function R_{sq} of a function H.

$$R_{\text{sq}}(H) = \frac{1}{n}\sum_{i=1}^{n}(H(x_i) - y_i)^2$$

- But since H is linear, we know $H(x) = w_1 x + w_0$.

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}((w_1 x_i + w_0) - y_i)^2$$

- Now MSE is a function of w_1, w_0.

Updated Goal

- Find slope w_1 and intercept w_0 which minimize the MSE, $R_{\text{sq}}(w_0, w_1)$:

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} \left((w_1 x + w_0) - y_i \right)^2$$

- Strategy: multivariable calculus.

Recall: the **gradient**

- If $f(x, y)$ is a function of two variables, the **gradient** of f at the point (x_0, y_0) is a **vector** of **partial derivatives**:

$$\nabla f(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) \\ \\ \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix}$$

- **Key Fact #1**: Derivative is to tangent line as gradient is to tangent plane.

- **Key Fact #2**: Gradient points in direction of biggest increase.

- **Key Fact #3**: Gradient is zero at critical points.

Strategy

To minimize $R(w_0, w_1)$: compute the gradient, set equal to zero, solve.

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)^2$$

Question

Choose the expression that equals $\frac{\partial R_{\text{sq}}}{\partial w_0}$.

a) $\frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)\, x_i$

b) $\frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)$

c) $\frac{2}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)\, x_i$

d) $\frac{2}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)$

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)^2$$

$$\frac{\partial R_{\text{sq}}}{\partial w_0} =$$

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)^2$$

Question

Choose the expression that equals $\frac{\partial R_{\text{sq}}}{\partial w_1}$.

a) $\frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)\, x_i$

b) $\frac{1}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)$

c) $\frac{2}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)\, x_i$

d) $\frac{2}{n} \sum_{i=1}^{n} ((w_1 x_i + w_0) - y_i)$

$$R_{\mathrm{sq}}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}\left((w_1 x_i + w_0) - y_i\right)^2$$

$$\frac{\partial R_{\mathrm{sq}}}{\partial w_1} =$$

Strategy

$$0 = \frac{2}{n}\sum_{i=1}^{n}\left((w_1 x_i + w_0) - y_i\right) \qquad 0 = \frac{2}{n}\sum_{i=1}^{n}\left((w_1 x_i + w_0) - y_i\right) x_i$$

1. Solve for w_0 in first equation.
2. Plug solution for w_0 into second equation, solve for w_1.

Solve for w_0

$$0 = \frac{2}{n} \sum_{i=1}^{n} \left((w_1 x_i + w_0) - y_i \right)$$

Solve for w_1

$$0 = \frac{2}{n} \sum_{i=1}^{n} \left((w_1 x_i + w_0) - y_i \right) x_i \qquad w_0 = \bar{y} - w_1 \bar{x}$$

Equivalent Formula for w_1

$$w_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})x_i}{\sum_{i=1}^{n}(x_i - \bar{x})x_i} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Key Fact

- Define

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- Then

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \sum_{i=1}^{n}(y_i - \bar{y}) = 0$$

Least Squares Solutions

- The **least squares solutions** for the slope w_1 and intercept w_0 are:

$$w_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad w_0 = \bar{y} - w_1\bar{x}$$

  where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- **Next Time:** We'll do an example and interpret these formulas.