Regression I

COMP9417, 22T2

Thinking Machine Learning

We try to make sense of data using mathematics to help us quantify what we *know*.

A standard way to break the problem down is as follows:

- We have 'input' data X and targets/outputs y
- Our data can be modelled as $y = f(X)$
- Goal is to find the best approximation for f as $\hat{f}$

We define the quality of our approximation ($\hat{f}$) by using a error/loss function.

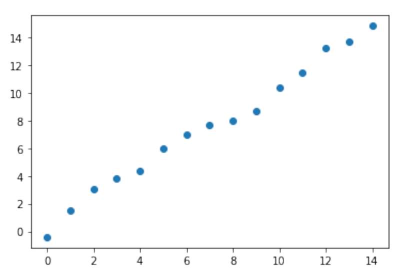
Linear Regression

We deduct and assume a linear relationship between X and y.

In this simple case, our model will take the form:

$$\hat{y} = w_0 + w_1 X$$

How do we find the optimal w_0 and w_1?



What will our loss function need? Boils down to the properties of the target function.

- Target function has ≈ 0 distance to all points

- We can define a basic loss function with one glaring issue:

$$L(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})$$

To make life easy, define a loss function:

$$\begin{aligned} L(w_0, w_1) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 && \text{a.k.a MSE} \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i)^2 && \text{by definition} \end{aligned}$$

The minimum of our loss function w.r.t w_0 and w_1 will be their optimal values respectively.

Question 1 (a, b, c)

1a

Derive the least-squares estimates for the univariate linear regression model.

i.e Solve:

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

First,

$$\begin{aligned} \frac{\partial L(w_0, w_1)}{\partial w_0} &= -\frac{2}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1 x_i) \\ &= -\frac{2}{n}\left(\sum_{i=1}^{n} y_i - n w_0 - w_1 \sum_{i=1}^{n} x_i\right) \end{aligned}$$

For the minimum, $\frac{\partial L(w_0,w_1)}{\partial w_0} = 0$,

$$-\frac{2}{n}\left(\sum_{i=1}^{n} y_i - nw_0 - w_1\sum_{i=1}^{n} x_i\right) = 0$$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} y_i - w_0 - w_1\frac{1}{n}\sum_{i=1}^{n} x_i &= 0 \\
\bar{y} - w_0 - w_1\bar{x} &= 0 \\
w_0 &= \bar{y} - w_1\bar{x}
\end{aligned} \tag{1}$$

To find w_1, we follow a similar process and use simple simultaneous equations to solve for the final solution.

So,

$$\begin{aligned}
\frac{\partial L(w_0,w_1)}{\partial w_1} &= -\frac{2}{n}\sum_{i=1}^{n} x_i(y_i - w_0 - w_1x_i) \\
&= -\frac{2}{n}\left(\sum_{i=1}^{n} x_iy_i - w_0\sum_{i=1}^{n} x_i - w_1\sum_{i=1}^{n} x_i^2\right)
\end{aligned}$$

$\frac{\partial L(w_0,w_1)}{\partial w_1} = 0$,

$$\begin{aligned}
\frac{1}{n}\left(\sum_{i=1}^{n} x_iy_i - w_0\sum_{i=1}^{n} x_i - w_1\sum_{i=1}^{n} x_i^2\right) &= 0 \\
\overline{xy} - w_0\bar{x} - w_1\overline{x^2} &= 0
\end{aligned}$$

$$\begin{aligned}
\overline{xy} - w_0\bar{x} - w_1\bar{x^2} &= 0 \\
w_1 &= \frac{\overline{xy} - w_0\bar{x}}{\overline{x^2}}
\end{aligned} \tag{2}$$

Sub (1) into (2):

$$\begin{aligned}
w_1 &= \frac{\overline{xy} - (\bar{y} - w_1\bar{x})\bar{x}}{\overline{x^2}} \\
w_1 &= \frac{\overline{xy} - \bar{x}\bar{y} + w_1\bar{x}^2}{\overline{x^2}} \\
w_1(\frac{\overline{x^2} - \bar{x}^2}{\bar{x}^2}) &= \frac{\overline{xy} - \bar{x}\bar{y} + w_1\bar{x}^2}{\overline{x^2}} \\
w_1 &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}
\end{aligned}$$

Finally, we have

$$w_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \text{ and } w_0 = \bar{y} - \bar{x}\frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$

1b

Problem: Prove $(\bar{x}, \bar{y})$ is on the line.

From 1(a), we have

$$\hat{y} = \bar{y} - \bar{x}\frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} + \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}x$$

Sub $x = \bar{x}$,

$$\begin{aligned}
\hat{y} &= \bar{y} - \bar{x}\frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} + \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}\bar{x} \\
\hat{y} &= \bar{y} \qquad \blacksquare
\end{aligned}$$

1c

Similar to 1a, though take care with the partial derivatives:

$$\begin{aligned}
\frac{\partial L(w_0, w_1)}{\partial w_0} &= -\frac{2}{n}\sum_{i=1}^{n}(y_i - w_0 - w_1x_i) \\
\frac{\partial L(w_0, w_1)}{\partial w_1} &= -\frac{2}{n}\sum_{i=1}^{n}x_i(y_i - w_0 - w_1x_i) + 2\lambda w_1
\end{aligned}$$

Final result is:

$$
\begin{aligned}
w_0 &= \bar{y} - w_1\bar{x} \\
w_1 &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2 + \lambda}
\end{aligned}
$$

Notice how the coefficients have an inverse relationship with λ.

Question 2 (a, b, c, d)

2a

Problem: Show that $\mathcal{L}(w) = \frac{1}{n}\|y - Xw\|_2^2$ has critical point $\hat{w} = (X^TX)^{-1}X^Ty$.

To find optimal w, solve $\dfrac{\partial \mathcal{L}(w)}{\partial w} = 0$

$$
\begin{aligned}
\mathcal{L}(w) &= \frac{1}{n}(y - Xw)^T(y - Xw) \\
&= \frac{1}{n}\left(y^Ty - y^TXw - w^TX^Ty + w^TX^TXw\right) \\
&= \frac{1}{n}\left(y^Ty - 2y^TXw + w^TX^TXw\right)
\end{aligned}
$$

$$
\frac{\partial \mathcal{L}(w)}{\partial w} = -\frac{1}{n}(-2X^ty + 2X^TXw)
$$

To solve for $\hat{w}$,

$$
\begin{aligned}
-2X^ty + 2X^TX\hat{w} &= 0 \\
\hat{w} &= (X^TX)^{-1}X^Ty
\end{aligned}
$$

2b

Problem: Prove $\hat{w} = (X^T X)^{-1} X^T y$ is a global minimum.

$$\begin{aligned}\nabla_w^2 \mathcal{L}(w) &= \nabla_w(\nabla_w \mathcal{L}(w)) \\ &= \nabla_w(-2X^T y + 2X^T X w) \\ &= 2X^T X\end{aligned}$$

So, for a vector $u \in \mathbb{R}^p$,

$$\begin{aligned}u^T(2X^T X)u &= 2(u^T X^T)(Xu) \\ &= 2(u^T X^T)(Xu) \\ &= 2(Xu)^T(Xu) \\ &= 2\|Xu\|_2^2 \geq 0\end{aligned}$$

Therefore, $\mathcal{L}$ is convex and $\hat{w}$ is the unique global minimum.

2c

2d

2e - Lab