Review via Linear Regression

02-680: Essentials of Mathematics and Statistics

December 5, 2024

The ***linear regression*** problem is as follows (we covered this briefly in Topic 10, but with a slightly different formula):

x — Independent variable[s] (what we control)

y — Dependent variable[s] (the measured values)

β_s – slope

β_0 – intercept

Some examples of regression in biology:

Independent	Dependent
Expression of genes expressed in immune response	asthma severity
Expression of regulator genets	Expression of target genes
Clinical variables	Insulin level
...	...

Lets assume we have some $x, y \in \mathbb{R}^n$, for n samples where each (x_i, y_i) are corresponding. Then

$$y = x^T \beta_s + \beta_0$$

To simplify things lets define: $X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_s \\ \beta_0 \end{bmatrix}$, then the equation above becomes simply:

$$y = X\beta.$$

Recall (possibly) from earlier that we want to find β that minimizes the $\|y - X\beta\|_2^2$, or the square of the L_2-norm.

$$
\begin{aligned}
\|y - X\beta\|_2^2 &= (y - X\beta)^T(y - X\beta) \\
&= (y^T - \beta^T X^T)(y - X\beta) \\
&= y^T y + \beta^T X^T X\beta - 2\beta^T X^T y
\end{aligned}
$$

To find the critical point of the curve (i.e. the minimum) we need

$$
\begin{aligned}
\frac{d}{d\beta} y^T y + \beta^T X^T X\beta - 2\beta^T X^T y &= 0 \\
\frac{d}{d\beta}\beta^T X^T X\beta^2 - \frac{d}{d\beta} 2\beta^T X^T y &= 0 \\
2X^T X\beta - 2X^T y &= 0 \\
2X^T X\beta &= 2X^T y \\
\beta &= X^T y (X^T X)^{-1}
\end{aligned}
$$

What if I want to know the probability of some y_i given x_i? (so $p(y_i \mid x_i)$)

Assuming its some natural process, we can assume

$$
p(y_i \mid x_i) \sim \mathcal{N}(\beta_0 + \beta_s x_i, \sigma^2).
$$

or equivalently

$$
y = \beta_0 + \beta_s x_i + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2).
$$

So given the data, we can actually use MLE to find β!

$$
\begin{aligned}
\ln p(\mathcal{D} \mid \beta, \sigma^2) &= \ln\left[\prod_{i=1}^{n} \frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2}\right] \\
&= \sum_{i=1}^{n} \ln\left[\frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2}\right] \\
&= \frac{-n}{2}\ln\sigma^2 + \sum_{i=1}^{n} \ln e^{-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2} \\
&= \frac{-n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \beta}\frac{-n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2 &= \frac{\partial}{\partial \beta} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2 \\
&= \frac{\partial}{\partial \beta} - \frac{1}{2\sigma^2}(Y - X\beta)^2 \\
&= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}(Y^2 - 2Y^TX\beta + \beta^TX^TX\beta) \\
&= ...
\end{aligned}
$$

$\hat{\beta}_{MLE} = X^Ty(X^TX)^{-1}$, thats just Least Squares!

But we don't know σ, thats actually, what we want to know because it models how uncertain we are with any given output.

$$
\frac{\partial}{\partial \sigma^2}\frac{-n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2 = \frac{-n}{2\sigma^2} + \frac{1}{2}(\sigma^2)^{-2}\sum_{i=1}^{n}(y_i - x_i\beta)^2
$$

$\hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^{n}(y_i - x_i\beta)^2}{n}$

With these we can determine for some new $\tilde{y}$ and $\tilde{x}$ what $p(\tilde{y} \mid \tilde{x})$ is.