

Linear Regression Models(draft)

Here are some materials I found interesting and compiled at the end of the course of Linear Regression Models.

To start with, the essence of linear regression models is actually conditional expectation. That's to say, given \mathbf{X} , the best estimator of \mathbf{Y} is $\mathbb{E}(\mathbf{Y}|\mathbf{X})$. And further, if we assume that $\mathbb{E}(\mathbf{Y}|\mathbf{X})$ is a linear function of \mathbf{X} , here comes the following interesting theorem.

Theorem 1

Suppose that \mathbf{X} and \mathbf{Y} have a joint distribution with means μ_X and μ_Y , standard deviations σ_X and σ_Y , and correlation ρ . Show that if $\mathbb{E}(\mathbf{Y}|\mathbf{X})$ is a linear function of \mathbf{X} , then

$$\mathbb{E}(\mathbf{Y}|\mathbf{X}) = \mu_Y + \rho\sigma_Y\left(\frac{\mathbf{X}-\mu_X}{\sigma_X}\right)$$

Its proof is not very difficult.

We may find this theorem is quite similar to the theorem below:

Theorem 2

Let $(\mathbf{X}_1, \mathbf{X}_2)$ have a joint p.d.f of the form below:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}}$$

The conditional distribution of \mathbf{X}_2 given $\mathbf{X}_1 = x_1$ is the normal distribution with mean and variance given by

$$\mathbb{E}(\mathbf{X}_2|\mathbf{X}_1 = x_1) = \mu_1 + \rho\sigma_2\left(\frac{x_1-\mu_1}{\sigma_1}\right), \text{Var}(\mathbf{X}_2|\mathbf{X}_1 = x_1) = (1 - \rho^2)\sigma_2^2$$

Without normal assumption

To some extent, we may relax the condition of the distribution of the error to not just normal but any well-defined distribution and still obtain the results under normality.

simple

multiple

With normal assumption

Most of the time, however, we are interested in the inference problem. And here comes into existence the assumption of normality.

simple

multiple

The independence between coefficients estimates and sum of squared residuals

VIF

VIF is an important index for the problem of multicollinearity. And we will give its derivation below which also illustrates why it can be taken as a sign of multicollinearity.

Just to mention, one approach to handle multicollinearity could be ridge regression.

Regularization

0-norm regularization

1-norm regularization

2-norm regularization

Under the condition of orthogonalization

Things become simple when the design matrix is column orthogonal. And we will talk about the behavior of coefficient estimators under this situation.

1-norm regularization

2-norm regularization

Diagnostic and Remedial

Actually, for me this is supposed to be the most important part which really requires lots of experience. Since I don't have much experience, however, I'm not gonna talk much about it. But maybe one day, I'll come back and finish this part. lol

The diagnostic is actually centered around the four assumptions, which are ...

While for the remedial,