

The Gaussian Distribution

Authors: Luke Bravo, Sidd Chauhan and Shuming Chen. (PDF)

The lecture covered the basics of multiple linear regression and the application of maximum likelihood estimation. Fundamental to both of these topics is the basic gaussian distribution, which these notes cover first. Then, we cover maximum likelihood functions and estimations. Next, we look at the equation, assumptions, and limitations of multiple linear regression. Finally, we have included two videos that provide more information on maximum likelihood estimation, and tie these three topics together.

Gaussian Distribution

Gaussian Distribution is interchangeable with normal probability distribution or a ‘bell curve’. It is symmetrical about the mean of its data points, and points closer to the mean are more common than points far off from the mean. A key characteristic of data that have a gaussian or normal distribution is that they are independent and identically distributed (i.i.d.). The i.i.d. assumption is central to machine learning theory; it asserts that data are consistent and uniform by virtue of the way they were sampled.

In general terms, functions with gaussian distributions are of the following form, where a and b are constants and c is a non-zero constant:

The equation below is one way of writing the probability density for the Gaussian distribution, where μ is the mean, σ is the standard deviation, and σ^2 is the variance.

The function has its peak at the mean, and its “spread” increases with the standard deviation (the function reaches 0.607 times its maximum at $x + \sigma$ and $x - \sigma$). This implies that normal is more likely to return samples lying close to the mean, rather than those far away. *From numpy documentation and P. R. Peebles Jr., “Central Limit Theorem” in “Probability, Random Variables and Random Signal Principles”, 4th ed., 2001, pp. 51, 51, 125.*

Transformation Of Variables To Follow Gaussian Distribution:

These are some direct, mathematical methods to transform variables so that they follow Gaussian distribution. None of them is better than the other. They

majorly depend on the original distribution of the variables. They are,

- Logarithmic Transformation
- Reciprocal Transformation
- Square Root Transformation
- Exponential Transformation

Maximum Likelihood Estimation

Maximum Likelihood Estimation is a method for estimating parameters of a probability distribution. We saw in class that maximizing a likelihood function (or minimizing its negative in the context of a cost function, see example equation below) lets you calculate the mean and variance of a distribution.

In most cases, it is best to take the natural log of a likelihood function and then maximize or minimize the result; this is then called log-likelihood. This simplifies multiplication to addition and cancels the e term. You can also avoid differentiating to test for maxima and then differentiating again by doing this. Professor Ghosh explained that we can rest assured that the extrema of the log-likelihood function will correspond to those of the original likelihood function because the logarithm is a monotonically decreasing function.

The two following equations express the likelihood and log-likelihood functions, respectively:

Note that for the likelihood function,

is a product of two single-variable functions.

Multiple Linear Regression

A linear regression model is a statistical model that's frequently used in data science. It's also one of the basic building blocks of machine learning! Multiple linear regression (MLR/multiple regression) is a statistical technique. It can use several variables to predict the outcome of a different variable. The goal of multiple regression is to model the linear relationship between your independent variables and your dependent variable. It looks at how multiple independent variables are related to a dependent variable.

Multiple Linear Regression is a statistical technique that models the relationship between a scalar response and one or more explanatory variables. It is often fitted using the Least Square approach. The general equation for a multiple linear regression is as follows:

There are two assumptions behind the model, and they are as follows:

1. The conditional mean of Y is linear in the X_i variables
2. The error terms ϵ are normally distributed, independent from each other, and have constant variance.

The model also has its limitations, one of them being that it does not model the distribution of x .

MLR is used extensively in econometrics and financial inference.

Minimizing the mean square error is equivalent to maximizing likelihood under a Gaussian model.

Least squares estimation finds the line that minuses total squared distance between the actual data point and the regression line. Maximum likelihood estimation maximizes the total probability of the data. Under Gaussian distribution, maximum probability is close to the mean value and this is equivalent to minimizing the distance between the data point and the mean value since Gaussian distribution is symmetric.

Resources

Gaussian Distribution

This article further explains gaussian distribution.

Maximum Likelihood Estimation

This video gives a brief overview of MLE. The example provided helps better visualize how the estimation works.

This article further explains MLE.

Multiple Linear Regression

This article further explains MLR and how to write the basic code for MLR.

This article explains how to apply MLR in investing and finance contexts.