Nathan Sponberg
Math 564

Project Proposal

Introduction.

The least squares method for generating a regression model is a powerful tool used in statistics, economics and the physical and social sciences. We briefly explore the derivation of this mathematical method and attempt to develop a basic intuition of the geometry involved in it. This leads to the realization that the least-squares method is in fact a projection operator in a real valued, finite-dimensional Hilbert space.

Least Squares.

To begin we will examine the derivation for a linear regression model using the least squares method. Suppose that we have a set of points $(x_1, y_1)...(x_n, y_n)$ that represent some set of observations. We would like to determine a linear equation that we can use to make predictions about future observations based on the points we have already observed. We assume the equation has the form $y = \beta_1 x + \beta_0$ where β_0 and β_1 are unknown. We determine the "best" values of these constant terms by minimizing the sum of the squared *vertical* distance for each of the observed points from the line $y = \beta_1 x + \beta_0$. Note that we denote these optimized values as $\hat{\beta}_0$ and $\hat{\beta}_1$. Using this method it can be shown that the following values of β_1 and β_0 minimize the sum of the squared distances:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\,.$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Where $\overline{x}$ and $\overline{y}$ are the means of the observed x and y values respectively.

Proof. For $i = 1, 2, ..., n$ the distance from the line to the observed value of y_i is given by

$$|y_i - \beta_1 x_i - \beta_0|\,.$$

The sum of the squared distances is then

$$S = \sum_{i=1}^{n}(y_i - \beta_1 x_i - \beta_0)^2\,.$$

We then calculate the partial derivatives of this expression in terms of β_1 and β_0 which we can then to use to find a minimum value. Observe that

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i(y_i - \beta_1 x_i - \beta_0)$$

and

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_1 x_i - \beta_0)\,.$$

Setting these two derivatives equal to zero we have

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \beta_1 x_i^2 + \beta_0 \sum_{i=1}^{n} x_i$$

and

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \beta_1 x_i + n\beta_0\,.$$

Solving for β_0 in the second equation yields

$$\beta_0 = \frac{\sum_{i=1}^n y_i - \beta_1 \sum_{i=1}^n x_i}{n} = \overline{y} - \beta_1 \overline{x}\,.$$

Note that this is the expression is optimized when $\beta_1 = \hat{\beta}_1$. Hence, $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$. We us this to solve for $\hat{\beta}_1$. Observe

$$\sum_{i=1}^n x_i y_i = \hat{\beta}_1 \sum_{i=1}^n x_i^2 + (\overline{y} - \hat{\beta}_1 \overline{x}) \sum_{i=1}^n x_i = \hat{\beta}_1 \sum_{i=1}^n x_i^2 + \overline{y} \sum_{i=1}^n x_i - \hat{\beta}_1 \overline{x} \sum_{i=1}^n x_i\,.$$

Rearranging terms and isolating $\hat{\beta}_1$ we obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \overline{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i}\,.$$

This expression can be rewritten in the desired form. First we consider the numerator, observe that

$$\sum_{i=1}^n x_i y_i - \overline{y} \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i - 2\overline{y} \sum_{i=1}^n x_i + \overline{y} \sum_{i=1}^n x_i =$$

$$\sum_{i=1}^n x_i y_i - 2\frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n} + \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n} = \,.$$

$$\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n} - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} + \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n}$$

We can then re-index the sums as follows

$$\sum_{j=1}^n x_j y_j - \sum_{j=1}^n \frac{y_j \sum_{i=1}^n x_i}{n} - \sum_{j=1}^n \frac{x_j \sum_{i=1}^n y_i}{n} + \sum_{j=1}^n \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n^2} =$$

$$\sum_{j=1}^n \left[x_j y_j - \frac{y_j \sum_{i=1}^n x_i}{n} - \frac{x_j \sum_{i=1}^n y_i}{n} + \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n^2} \right] =$$

$$\sum_{j=1}^n (x_j y_j - y_j \bar{x} - x_j \bar{y} + \bar{y}\bar{x}) = \sum_{j=1}^n (x_j - \overline{x})(y_j - \bar{y})\,.$$

Replacing y_i with x_i we follows directly that the following equality also holds

$$\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i = \sum_{i=1}^n (x_i - \overline{x})^2\,.$$

Hence we have shown that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \overline{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i} = \frac{\sum_{j=1}^n (x_j - \overline{x})(y_j - \bar{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}\,.$$

As the final step we will show that this is in fact a minimum. Computing the second partial derivatives of S yields

$$\frac{\partial^2 S}{\partial \beta_1^2} = 2 \sum_{i=1}^n x_i^2$$

$$\frac{\partial^2 S}{\partial \beta_0^2} = 2n$$

and

$$\frac{\partial^2 S}{\partial \beta_1 \beta_0} = 2\sum_{i=1}^{n} x_i \,.$$

Note that

$$\frac{\partial^2 S}{\partial \beta_1^2}\frac{\partial^2 S}{\partial \beta_0^2} - \Big(\frac{\partial^2 S}{\partial \beta_1 \beta_0}\Big)^2 = 4n\sum_{i=1}^{n} x_i^2 - 4\Big(\sum_{i=1}^{n} x_i\Big)^2 > 0\,.$$

Therefore

$$\frac{\partial^2 S}{\partial \beta_1^2} \geq 0$$

for all β_1. Hence, it follows that the values of $\hat{\beta}_1$ and $\hat{\beta}_0$ calculated above give the minimum value of the sum of squares.

□

Extensions of Least Squares. A similar method as described above can be used to fit regression curves to higher degree polynomials in x_i that are of the form

$$x_i^n c_n + \cdots + x_i^2 c_2 + x_i\beta_1 + \beta_0\,.$$

A system of equations can be created by taking the partial derivatives with respect to the c_i. This system can then be solved (provided a solution exists) to find the minimum values for the c_i. It is also natural to extend this method to the multivariate case given a set points $(y_1, x_1^1, ..., x_n^1), (y_2, x_2^1, ..., x_2^n), ..., (y_m, x_m^1, ..., x_m^n)$. The technique is similar to the ones laid out above.

It may also be of interest and worth to examine how this method extends into different spaces. It maybe possible to use a similar technique in order to compute a least squares equation in a normed vector space with an arbitrary norm. Perhaps of particular interest would be a norm other than the standard euclidean distance used in the least squares regression described above.

Another area that maybe even more compelling to consider a least squares method, or a similar method of minimizing distance, that focuses on orthogonal distance as opposed to vertical distance. It maybe a natural extension of this method to consider developing an approached of minimizing the sum of orthogonal "distances" of a set of vectors in an arbitrary (probably finite dimensional) Hilbert space. Using the inner product of a given Hilbert space and a set of vectors in the space, one maybe able to find a "minimum" subspace that minimizes the orthogonal projects of the set of vectors.

These ideas bear further consideration and we propose to investigate what theory there is (if any) exploring these extensions of the least squares method into more general spaces.