Mathematical Statistics - Section 1 - NYU Spring 2019 - Homework 8

Quadratic regression We study the following problem: we have n data points $(X_1, Y_1), \dots, (X_n, Y_n)$ and we want to find the best *quadratic* fit. In other words, we are looking for coefficients a, b, c such that the fitted values

$$\hat{Y}_i := a + bX_i + cX_i^2$$

approximate the real values Y_i as well as possible. In which sense? Once again, we want the "residual sum of squares", i.e. the quantity

$$\sum_{i=1}^{n} |Y_i - \hat{Y}_i|^2$$

to be as small as possible. We will follow the same strategies as for the linear case.

Questions are written with a **bold** typeface.

1. First, we re-formulate the problem as follows: we let $\overrightarrow{U}$ be the vector with n coordinates equal to 1, $\overrightarrow{X}$ be the vector with coordinates $X_1, \dots, X_n$ and $\overrightarrow{X^2}$ be the vector with coordinates $X_1^2, \dots, X_n^2$. Let also $\overrightarrow{Y}$ be the vector with coordinates $Y_1, \dots, Y_n$.

   Find the orthogonal projection of $\overrightarrow{Y}$ onto the linear subspace spanned by $\overrightarrow{U}, \overrightarrow{X}, \overrightarrow{X^2}$. That is, find (equations characterizing) a, b, c such that the vector
   $$a\overrightarrow{U} + b\overrightarrow{X} + c\overrightarrow{X^2} - \overrightarrow{Y}$$
   is orthogonal to $\overrightarrow{U}, \overrightarrow{X}$ and $\overrightarrow{X^2}$.

   Write this as as linear system of three equations with unknowns a, b, c (if you're brave, you can solve it).

2. Next, we encode all the "features" in a matrix. First of all, we perform the trick of adding a 1 as the first coordinate of each feature, and for every $i = 1 \dots n$ we let

   $$\text{Feature}_i := (1, X_i, X_i^2),$$

   and then we let F be the matrix whose rows are the Feature_i for $i = 1 \dots n$, which means

   $$F = \begin{pmatrix} 1 & X_1 & X_1^2 \\ \dots & \dots & \dots \\ 1 & X_n & X_n^2 \end{pmatrix}$$

By construction, F is an "n by 3" matrix. We ask the following question: which element in the image of F, i.e. which element of the form FC (where $C = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is a vector with three arbitrary coordinates a, b, c), is the closest to $L = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}$? Once again, it is the projection of L onto the correct space, here the image of F.

Show that the vectors in the image of F are the same thing as the vectors of the form $a\overrightarrow{U} + b\overrightarrow{X} + c\overrightarrow{X^2}$ considered above.

We have thus already "solved" this problem in question 1.

Solve it a second time in a slightly more "abstract" way, as done in class for the "multiple linear regression".