September 13, 2022

1

By assuming that the continous funcitiion $f(x)$ can be parametrized in terms of a polynomial of degree $n-1$ means that our datapoints can be described by the following expression:

$$\boldsymbol{y} = \sum_{j=0}^{n-1} \beta_j x_i^j + \epsilon_i$$

This is the same as the following matrix equation:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} = \boldsymbol{\tilde{y}} + \boldsymbol{\epsilon}$$

This means that the elemnt i is given by:

$$y_i = \epsilon_i + \sum_j x_{ij}\beta_j$$

The expactation value of the element i in $\boldsymbol{y}$:

$$E(y_i) = E(\epsilon_i + \sum_j x_{ij}\beta_j)$$

Since we already know that $\boldsymbol{\epsilon}$ is normal distibuted with the expectation value 0 $E(\epsilon_i) = 0$. This gives us:

$$E(y_i) = E(\epsilon_i) + E(\sum_j x_{ij}\beta_j) = \sum_j E(x_{ij}\beta_j)$$

x_{ij} and β_j are constants which have themselves as expectation value leaves us with:

$$E(y_i) = \sum_j x_{ij}\beta_j = \boldsymbol{X_{i*}\beta}$$

To find the variance we again recognize that $\boldsymbol{X\beta}$ follow no distribution which leaves us with only the variance of $\boldsymbol{\epsilon}$ which is given as σ^2:

$$\mathrm{Var}(\boldsymbol{y}) = \mathrm{Var}(\boldsymbol{X\beta} + \boldsymbol{\epsilon}) = \mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2$$

$\tilde{\boldsymbol{y}}$ is defined through the minimization of the mean square error $(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2$ which for $\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$ translates to the minimization of the cost function:

$$C(\boldsymbol{\beta}) = \frac{1}{n}\{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^T\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\}$$

Where we take the derivative with respect to $\boldsymbol{\beta}$ and solve where the derivative is 0 to find the minimum:

$$\frac{\partial C(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) = 0$$

$$\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}$$

We assume that $\boldsymbol{X}^T\boldsymbol{X}$ is invertible which gives us the the optimal $\boldsymbol{\beta}$:

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

We can now find the expectation value for the optimal $\tilde{\boldsymbol{\beta}}$:

$$\begin{aligned}
E(\tilde{\boldsymbol{\beta}}) &= E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}] = E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E[(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}$$

Here we see that the expectation value of our optimal parameter is the parameter $\boldsymbol{\beta}$.

We now find the variance of the optimal parameter:

$$\begin{aligned}
\mathrm{Var}(\tilde{\boldsymbol{\beta}}) &= \mathrm{Var}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})] \\
&= \mathrm{Var}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\epsilon}]
\end{aligned}$$

For a matrix A we have $\mathrm{Var}(\boldsymbol{A}\boldsymbol{X}+b) = \boldsymbol{A}\mathrm{Var}(\boldsymbol{X})\boldsymbol{A}^T$ which gives us:

$$\begin{aligned}
\mathrm{Var}(\tilde{\boldsymbol{\beta}}) &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{Var}[\boldsymbol{\epsilon}]((\boldsymbol{X}^T\boldsymbol{X})^{-1})^T\boldsymbol{X} \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{Var}[\boldsymbol{\epsilon}](\boldsymbol{X}^T)^{-1}\boldsymbol{X}^{-1}\boldsymbol{X} \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\mathrm{Var}[\boldsymbol{\epsilon}]
\end{aligned}$$

2