1 Linear Regression

Setup:

- Start with a random vector $X^T = (X_1, \cdots X_p)$ and random variable Y.
- Will approximate regression function $Y = f(X)$ where $f(x) = E(Y|X = x)$ with linear function $f(x) = \sum_{j=1}^{p} \beta_j X_j$
- let $\mathbf{X}$ be the $n \times p$ matrix of observations of X, and y be the observations of Y.
- easy to show that minimizng squared error loss function yield the following estimate $\hat{\beta}$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$$

- now, from ESL book: "In order to pin down the sampling properties of $\hat{\beta}$, we now assume that the observations y_i are uncorrelated and have constant variance σ^2, and that the [observations] x_i are fixed (non-random)"

Results:

- first, from solution for $\hat{\beta}$ we can prove that

$$VAR(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-}1\sigma^2$$

This is relatively straight forward. Can prove $VAR(AZ) = AVAR(Z)A^T$ for constant matrix A and random vector Z. And uncorrelated, constant variance implies $VAR(Y) = \mathbf{I_n}\sigma^2$.

- second: the typical estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{N-p-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2.$$

This estimator is unbiased - that is: $E(\hat{\sigma}^2) = \sigma^2$.

I cannot prove this. But ESL simply brush past it, because I assume it is a very well known (and perhaps very easy) result. This is what I need help with.

- Then they say: "To draw inferences about the parameters and the model, additional assumptions are needed" - they asssume:
  - the conditional expectation of Y really is linear in the X
  - the deviations of Y around its expectation are additive and gaussian, hence

$$Y = E(Y|X_1, \cdots X_p) + \epsilon \tag{1}$$

    where $\epsilon \sim N(0, \sigma^2)$

  Now it is easy to show

$$\hat{\beta} \sim N(\beta, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$$

  to do this, think of each observation y_i as a random variable Y_i, which by assumption is $Y_i = E(Y|X = x_i) + \epsilon_i$, where the ϵ_i are i.i.d $N(0, \sigma^2)$. Plugging this in, yields said result.