Project 1

Part a)

Assuming that there exist a cont. function $f(\vec{x})$ and normal distr. of error $\varepsilon \sim N(0, \sigma^2)$

This describes our data

$$\vec{y} = f(\vec{x}) + \varepsilon = \tilde{\vec{y}} + \varepsilon = X\beta + \varepsilon$$

a.1) Show $\mathbb{E}(y_i) = \sum_j x_{ij}\beta_j = X_{i,*}\beta$

$$\mathbb{E}(y_i) = \langle y_i \rangle = \langle \sum_j x_{ij}\beta_j + \varepsilon_i \rangle$$

Because of linearity, we have $\langle \sum_j x_{ij}\beta_j \rangle + \langle \varepsilon \rangle$

We see that term 1 sum over the whole row on ith column. Meaning the content is a scalar, and the expected value of a scalar is a scalar itself.

As for the second term, $\langle \varepsilon \rangle$, we have assumed that $\varepsilon \sim N(0, \sigma^2) \Rightarrow \langle \varepsilon \rangle = 0$

$$\Rightarrow \langle y_i \rangle = \sum_j x_{ij}\beta_j$$

a.2) Show $\mathrm{Var}(y_i) = \sigma^2$

Variance is a measurement of how far each number in data (y_i) is "away" from mean$(\mathbb{E}(y_i))$.

$\langle (y_i - \langle y_i \rangle)^2 \rangle$ again, because of linearity:

$$= \langle y_i^2 - 2y_i\langle y_i \rangle + \langle y_i \rangle^2 \rangle = \langle y_i^2 \rangle - \langle 2y_i \langle y_i \rangle \rangle + \langle \langle y_i \rangle^2 \rangle$$

$$= \langle y_i^2 \rangle - \langle y_i \rangle \langle 2y_i \rangle + \langle y_i \rangle^2$$

$$= \langle y_i^2 \rangle - 2\langle y_i \rangle \langle y_i \rangle + \langle y_i \rangle^2 = \langle y_i^2 \rangle - 2\langle y_i \rangle^2 + \langle y_i \rangle^2$$

$$= \langle y_i^2 \rangle - \langle y_i \rangle^2$$

We know that $y_i = \sum_j x_{ij}\beta_j + \varepsilon_i$

$$\Rightarrow \langle (\sum_j x_{ij}\beta_j + \varepsilon_i)^2 \rangle - (\sum_j x_{ij}\beta_j + \cancel{\varepsilon_i})^2$$

$$= \langle (\sum_j x_{ij}\beta_j)^2 + 2\sum_j x_{ij}\beta_j\varepsilon_i + \varepsilon_i^2 \rangle - (\sum_j x_{ij}\beta_j + \cancel{\varepsilon_i})^2$$

$$= \langle (\sum_j \cancel{x_{ij}}\beta_j)^2 \rangle + 2\sum_j x_{ij}\beta_j \underbrace{\langle \varepsilon_i \rangle}_{0} + \langle \varepsilon_i^2 \rangle - (\sum_j \cancel{x_{ij}}\beta_j)^2 = \langle \varepsilon_i^2 \rangle = \sigma^2$$