$$
\begin{aligned}
E(y_i) &= E(f(x_i) + \epsilon) \\
&= E(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_{n-1} x_{in-1} + \epsilon) \\
&= E\left(\sum_{j=0}^{n-1} \beta_j x_{ij}\right) + E(\epsilon) \\
&= \sum_{j=0}^{n-1} \beta_j x_{ij}
\end{aligned}
$$

Where we've used the linearity of expectation.

$$
\begin{aligned}
V(y_i) &= V(f(x_i) + \epsilon) \\
&= V(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_{n-1} x_{in-1} + \epsilon) \\
&= V(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_{n-1} x_{in-1}) + V(\epsilon) \\
&= V(\epsilon) \\
&= \sigma^2
\end{aligned}
$$

Where we've used that the variance is invariant.

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

Where we've again used the linearity of the expectation.

$$
\begin{aligned}
V(\hat{\beta}) &= V[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}] \\
&= [(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]\; V[\boldsymbol{y}]\; [(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]^T \\
&= [(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]\; \sigma^2\boldsymbol{I}\; [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}] \\
&= \sigma^2[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}] \\
&= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}
\end{aligned}
$$

Where we've used that if $\boldsymbol{A}$ is a contant matrix, $V(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}V(\boldsymbol{X})\boldsymbol{A}^T$. **e)**
Derivation of the bias-variance tradeoff:

$$
\begin{aligned}
E\left[(\boldsymbol{y} - \tilde{\boldsymbol{y}})^2\right] &= E\left[(\boldsymbol{y} - E(\tilde{\boldsymbol{y}}) + E(\tilde{\boldsymbol{y}}) - \tilde{\boldsymbol{y}})^2\right] \\
&= E\left[(\boldsymbol{y} - E(\tilde{\boldsymbol{y}}))^2 + 2(\boldsymbol{y} - E(\tilde{\boldsymbol{y}}))(E(\tilde{\boldsymbol{y}}) - \tilde{\boldsymbol{y}}) + (\tilde{\boldsymbol{y}} - E(\tilde{\boldsymbol{y}}))^2\right] \\
&= E\left[(\boldsymbol{y} - E(\tilde{\boldsymbol{y}}))^2\right] + 2E\left[(\boldsymbol{y} - E(\tilde{\boldsymbol{y}}))(E(\tilde{\boldsymbol{y}}) - \tilde{\boldsymbol{y}})\right] + E\left[(\tilde{\boldsymbol{y}} - E(\tilde{\boldsymbol{y}}))^2\right]
\end{aligned}
$$

The first term in the expression above can be simplified as follows:

$$\begin{aligned}
E\left[(\boldsymbol{y}-E(\tilde{\boldsymbol{y}}))^2\right] &= E\left[(f(x)+\epsilon-E(\tilde{\boldsymbol{y}}))^2\right] \\
&= E\left[(f(x)-\tilde{\boldsymbol{y}})^2-2(\epsilon(E(\tilde{\boldsymbol{y}})-f(x)))+\epsilon^2\right] \\
&= E\left[(f(x)-\tilde{\boldsymbol{y}})^2\right]-2E\left[\epsilon(E(\tilde{\boldsymbol{y}})-f(x))\right]+E[\epsilon^2] \\
&= E\left[(f(x)-\tilde{\boldsymbol{y}})^2\right]-2E[\epsilon]E\left[E(\tilde{\boldsymbol{y}})-f(x)\right]+\sigma^2 \\
&= E\left[(f(x)-\tilde{\boldsymbol{y}})^2\right]+\sigma^2
\end{aligned}$$

If we now substitute this expression in () we get:

$$\begin{aligned}
E\left[(\boldsymbol{y}-\tilde{\boldsymbol{y}})^2\right] &= (E\left[f(x)-(\tilde{\boldsymbol{y}})^2\right]+\sigma^2)+2(E(\tilde{\boldsymbol{y}})-\boldsymbol{y})E\left[(\tilde{\boldsymbol{y}}-E(\tilde{\boldsymbol{y}}))\right]+V(\tilde{\boldsymbol{y}}) \\
&= Bias[\tilde{\boldsymbol{y}}]^2+2(E(\tilde{\boldsymbol{y}})-\boldsymbol{y})(E(\tilde{\boldsymbol{y}})-E(\tilde{\boldsymbol{y}}))+V(\tilde{y})+\sigma^2 \\
&= Bias[\tilde{\boldsymbol{y}}]^2+V(\tilde{y})+\sigma^2
\end{aligned}$$