Ex. 2.9 (the average training error is smaller than the testing error)

The expectation of the test term $\frac{1}{M}\sum\left(\tilde{y}_i - \hat{\beta}^T x_i\right)^2$ is equal to the expectation of $\left(\tilde{y}_1 - \hat{\beta}^T x_1\right)^2$, and is therefore independent of M. We take $M = N$, and then decrease the test expression on replacing $\hat{\beta}$ with a value of β that minimizes the expression. Now the expectations of the two terms are equal. This proves the result. Note that we may have to use the Moore-Penrose pseudo-inverse of $X^T X$, if the rank of X is less than p. This is not a continuous function of X, but it is measurable, which is all we need.