Effect of smaller training set

1		n
X_1	$\cdots$	X_n

Training set

1		m
Y_1	$\cdots$	Y_m

test/validation set

$X_i, Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$

Interested in the prediction error for $\{Y_i\}_{i=1}^{m}$

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \hat{Y}_i)^2$$

0^{th}-order prediction for Y_i using intercept

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

$$\mathbb{E}[MSE] = \mathbb{E}\left[\frac{1}{m}\sum_i (Y_i - \hat{\mu})^2\right] = \frac{1}{m}\sum_i \mathbb{E}[(Y_i - \hat{\mu})^2]$$

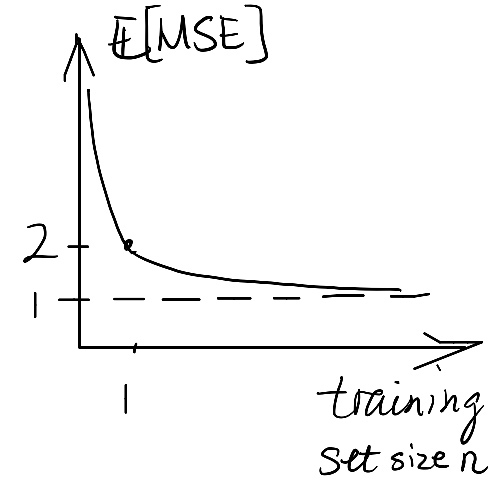
$$= \frac{1}{m}\sum_i \mathbb{E}\left\{\left[(Y_i - \mu) - (\hat{\mu} - \mu)\right]^2\right\}$$

$$= \frac{1}{m}\sum_i \left\{\mathbb{E}[(Y_i - \mu)^2] + \mathbb{E}[(\hat{\mu} - \mu)^2] + 2\underbrace{\mathbb{E}[(Y_i - \mu)(\hat{\mu} - \mu)]}_{0}\right\}$$

$$= \frac{1}{m}\sum_i \left\{Var(Y_i) + Var(\hat{\mu})\right\}$$

$$= \frac{1}{m}\sum_i \left(\sigma^2 + \frac{1}{n}\sigma^2\right)$$

$$= \left(1 + \frac{1}{n}\right)\sigma^2$$



As training set shrinks in size, prediction error increases.