There is a trade-off between bias and variance

$$\underbrace{\mathbb{E}\left(y - \hat{f}(x)\right)^2}_{\text{Expected } MSE \text{ in test set}} = \underbrace{\mathbb{V}\left(\hat{f}(x)\right)}_{\text{Variance}} + \underbrace{\left[\text{Bias}\left(\hat{f}(x)\right)\right]^2}_{\text{Squared Bias}} + \underbrace{\mathbb{V}(\epsilon)}_{\text{Irreducible Error}}$$

There is a trade-off between bias and variance

$$\underbrace{\mathbb{E}\left(y - \hat{f}(x)\right)^2}_{\text{Expected } MSE \text{ in test set}} = \underbrace{\mathbb{V}\left(\hat{f}(x)\right)}_{\text{Variance}} + \underbrace{\left[\text{Bias}\left(\hat{f}(x)\right)\right]^2}_{\text{Squared Bias}} + \underbrace{\mathbb{V}(\epsilon)}_{\text{Irreducible Error}}$$

- A model with high bias is a poor approximation of reality

There is a trade-off between bias and variance

$$\underbrace{\mathbb{E}\left(y - \hat{f}(x)\right)^2}_{\text{Expected } MSE \text{ in test set}} = \underbrace{\mathbb{V}\left(\hat{f}(x)\right)}_{\text{Variance}} + \underbrace{\left[\text{Bias}\left(\hat{f}(x)\right)\right]^2}_{\text{Squared Bias}} + \underbrace{\mathbb{V}\left(\epsilon\right)}_{\text{Irreducible Error}}$$

- A model with high bias is a poor approximation of reality
- A model with high variance is one that does not generalize well to a new data set

There is a trade-off between bias and variance

$$\underbrace{\mathbb{E}\left(y-\hat{f}(x)\right)^2}_{\text{Expected } MSE \text{ in test set}} = \underbrace{\mathbb{V}\left(\hat{f}(x)\right)}_{\text{Variance}} + \underbrace{\left[\text{Bias}\left(\hat{f}(x)\right)\right]^2}_{\text{Squared Bias}} + \underbrace{\mathbb{V}\left(\epsilon\right)}_{\text{Irreducible Error}}$$

- A model with high bias is a poor approximation of reality
- A model with high variance is one that does not generalize well to a new data set
- A model is overfit if it has low bias and high variance

There is a trade-off between bias and variance

$$\underbrace{\mathbb{E}\left(y-\hat{f}(x)\right)^2}_{\text{Expected } MSE \text{ in test set}} = \underbrace{\mathbb{V}\left(\hat{f}(x)\right)}_{\text{Variance}} + \underbrace{\left[\text{Bias}\left(\hat{f}(x)\right)\right]^2}_{\text{Squared Bias}} + \underbrace{\mathbb{V}\left(\epsilon\right)}_{\text{Irreducible Error}}$$

- A model with high bias is a poor approximation of reality
- A model with high variance is one that does not generalize well to a new data set
- A model is overfit if it has low bias and high variance
- A model is underfit if it has high bias and low variance

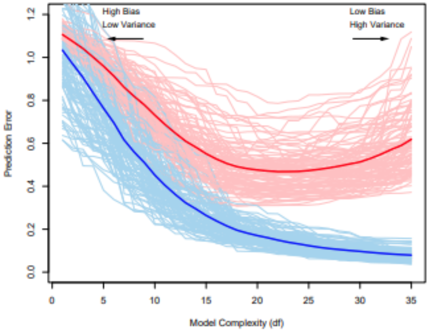

FIGURE 7.1. *Behavior of test sample and training sample error as the model complexity is varied. The light blue curves show the training error* $\overline{\mathrm{err}}$*, while the light red curves show the conditional test error* $\mathrm{Err}_{\mathcal{T}}$ *for* 100 *training sets of size* 50 *each, as the model complexity is increased. The solid curves show the expected test error* Err *and the expected training error* $\mathrm{E}[\overline{\mathrm{err}}]$.

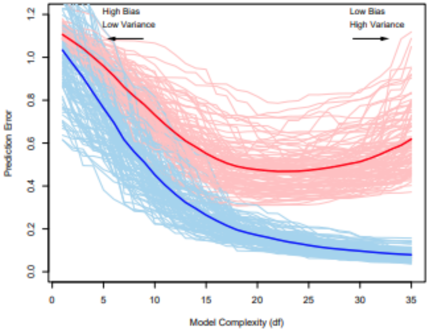

FIGURE 7.1. *Behavior of test sample and training sample error as the model complexity is varied. The light blue curves show the training error* $\overline{\text{err}}$*, while the light red curves show the conditional test error* $\text{Err}_{\mathcal{T}}$ *for 100 training sets of size 50 each, as the model complexity is increased. The solid curves show the expected test error* Err *and the expected training error* $\text{E}[\overline{\text{err}}]$.

- Optimal model complexity is at the minimum of the red line

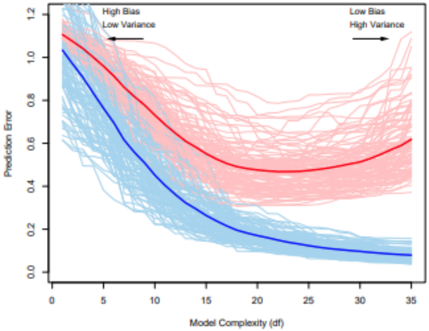

FIGURE 7.1. *Behavior of test sample and training sample error as the model complexity is varied. The light blue curves show the training error* $\overline{\mathrm{err}}$*, while the light red curves show the conditional test error* $\mathrm{Err}_{\mathcal{T}}$ *for* 100 *training sets of size* 50 *each, as the model complexity is increased. The solid curves show the expected test error* Err *and the expected training error* $\mathrm{E}[\overline{\mathrm{err}}]$.

- Optimal model complexity is at the minimum of the red line
- Irreducible error means that the red line can't go below a certain level

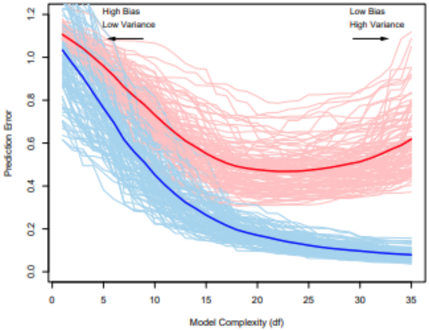

FIGURE 7.1. *Behavior of test sample and training sample error as the model complexity is varied. The light blue curves show the training error* $\overline{\text{err}}$*, while the light red curves show the conditional test error* $\text{Err}_{\mathcal{T}}$ *for 100 training sets of size 50 each, as the model complexity is increased. The solid curves show the expected test error* Err *and the expected training error* $\text{E}[\overline{\text{err}}]$.

- Optimal model complexity is at the minimum of the red line
- Irreducible error means that the red line can't go below a certain level
- Maybe at 0.3 in this picture?

When the target variable is continuous, we use MSE to measure fit

When the target variable is continuous, we use MSE to measure fit

When it is categorical, we use accuracy or similar measures

When the target variable is continuous, we use *MSE* to measure fit

When it is categorical, we use accuracy or similar measures

- Or some combination of specificity and sensitivity

When the target variable is continuous, we use *MSE* to measure fit

When it is categorical, we use accuracy or similar measures

- Or some combination of specificity and sensitivity
- Goal is to not have a good "fit" by randomly guessing

When the target variable is continuous, we use *MSE* to measure fit

When it is categorical, we use accuracy or similar measures

- Or some combination of specificity and sensitivity
- Goal is to not have a good "fit" by randomly guessing
- So each potential metric penalizes random guessing