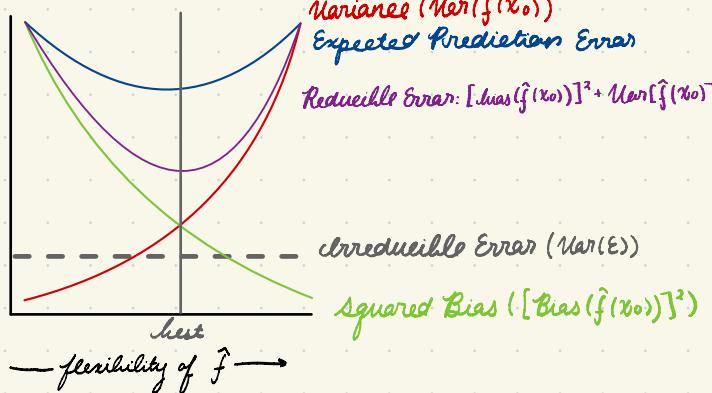


Homework I

I) a)



The ideal flexibility of a model minimizes the reducible error. Reducible error is minimized only when there is a balance between bias & variance. When variance is high, the model will be very flexible and fit the training data exactly, causing a very low bias. This overfitting will cause the model to be unable to adapt to new data, causing a low testing accuracy. Similarly, if the model has very low variance, the bias will become extremely high because it can't adapt to new examples, again causing a high error rate. The ideal flexibility is somewhere in between, allowing the model to minimize error.

- b)
 - a) Linear regression with just an intercept would be expected to have a lower variance & higher bias. Linear regression with a single predictor will increase the variance of the model, increasing its flexibility & decreasing bias.
 - b) A model with an MSE of 200 will have a higher bias & lower var than that of a model with an MSE of 100. A low training MSE indicates the model is fitting the data well, which is associated with a higher variance, leading to lower bias.
 - c) A model that interpolates will be expected to have higher variance & lower bias because it will pass through every known point, i.e. has a lower bias & higher variance.

c) No, given these two MSE's alone we cannot infer the model with greater variance/bias. MSE alone tells us the total error, not how this is distributed over the squared bias, variance, and irreducible error.

III) a) An example of a prediction task could be if certain levels of the covariate allow you to predict whether the outcome is T/F. You might report the level of the covariate where the prediction switches from T→F and the accuracy (MSE) at which the model is correct.

b) An example of an inference task using this data could be investigating whether the level of antibodies is associated with a reduced risk in infection. You could report odds-ratios or regression coefficients