Q. Bias Variance tradeoff

Training set :

True relationship → Bais

Squbly relation ship has less bias

Difference between different dataset is called variance (test/train) Overfitting

High variance : where the error is

low variance

Low variance: : (error is similar), consistent

Regularization, boosting and bagging helps in this

Upleveling : under parameterized vs over parameterized

How do you choose a model?

Ans) Bias vs Variance tradeoff

High Bias. low variance : consistent but inaccurate on average, [simple model[

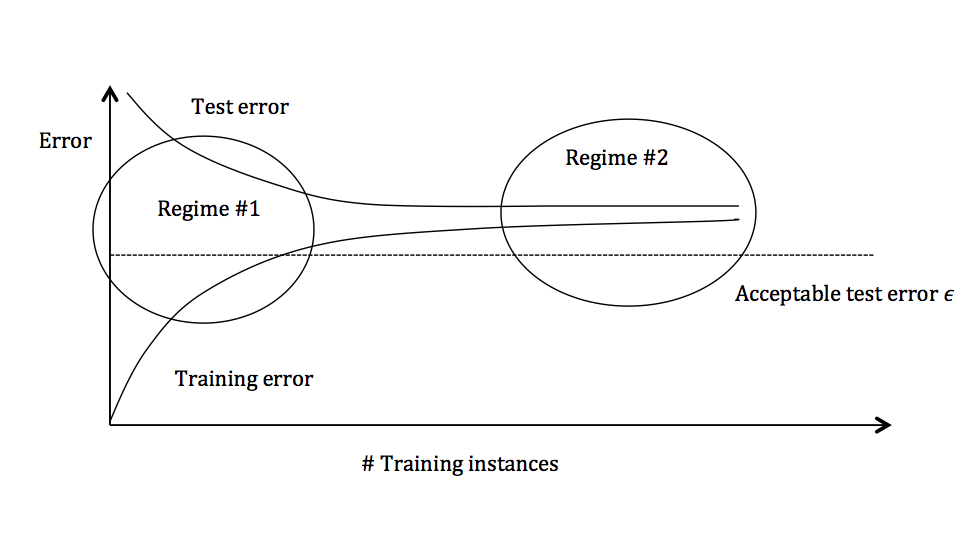
Low Bias, High variance: accurate on average but inconsistent [ complex model]

### **Detecting High Bias and High Variance**

**Variance**: Captures how much your classifier changes if you train on a different training set. How "over-specialized" is your classifier to a particular training set (overfitting)? If we have the best possible model for our training data, how far off are we from the average classifier?

**Bias**: What is the inherent error that you obtain from your classifier even with infinite training data? This is due to your classifier being "biased" to a particular kind of solution (e.g. linear classifier). In other words, bias is inherent to your model.

**Noise**: How big is the data-intrinsic noise? This error measures ambiguity due to your data distribution and feature representation. You can never beat this, it is an aspect of the data.



#### **Regime 1 (High Variance)**

In the first regime, the cause of the poor performance is high variance.

**Symptoms**:

1. Training error is much lower than test error
2. Training error is lower than ϵ
3. Test error is above ϵ

**Remedies**:

* Add more training data
* Reduce model complexity -- complex models are prone to high variance
* Bagging (will be covered later in the course)

#### **Regime 2 (High Bias)**

Unlike the first regime, the second regime indicates high bias: the model being used is not robust enough to produce an accurate prediction.

**Symptoms**:

1. Training error is higher thanϵ

**Remedies**:

* Use more complex model (e.g. kernelize, use non-linear models)
* Add features
* Boosting (will be covered later in the course)

—---------------------------------------

Regularization takes many flavors depending on the algorithm in question, but the common thread is that regularization introduces a small amount of bias in exchange for a reduction in variance

Regularizarion

1. Lasso - helps in reducing the variance (reducing overfitting by simplifying the model), job of penalizing only the high coefficients, does feature selection
2. Ridge - sum of square, doesnt bring it zero but subststialy small, all the features are used

To wrap things up, we can relate the Bias Variance decomposition to the

commonly used terms overfitting and underfitting in the following informal

way:

• Overfitting relates to having a High Variance model or estimator. To

fight overfitting, we need to focus on reducing the Variance of the estimator, such as: increase regularization, obtain larger data set, decrease

number of features, use a smaller model, etc.

• Underfitting relates to having a High Bias model or estimator. To fight

underfitting, we need to focus on reducing the Bias in the estimator,

such as: decrease regularization, use more features, use a larger model,

etc.

The first step in improving generalization error is to characterize which component in the decomposition has the highest contribution, and go after that

component. Unfortunately there is no theoretically sound yet tractable way

of calculating the breakdown. However there are certain heuristics that are

extremely useful. Loosely speaking:

• Training error can be treated as the amount of Bias in the model or

estimator. If the model is unable to fit the training data itself well, then

it is likely that the model has High Bias. This is the underfitting regime.

8

• Gap between cross-validation error and Training error can be treated as

the Variance of the model or the estimator. If the Training error is low

but the Cross Validation error is high, it is very likely that model has

High Variance. This is the overfitting regime.

We should always analyze the model performance by looking at the training error and cross-validation error simultaneously. This is the only tractable

(albeit heuristic) way to obtain an estimate of the Bias and Variance components. Only then should we take steps that are targeted towards addressing

either Bias or Variance purposefully.

Steps taken to fight overfitting (i.e. fight High Variance) generally do not

necessarily help fight underfitting (i.e. High Bias). For example, it is futile

to spend time and resources in obtaining more data (technique to fight High

Variance) when the training error itself is high (symptom of High Bias).

Similarly steps taken to fight underfitting (i.e. fight High Bias) generally

do not necessarily help fight overfitting (i.e. High Variance). For example,

it is futile to switch to a larger neural network (technique to fight High

Bias) when the gap between cross-validation error and training error is high

(symptom of High Variance).

Many times steps taken to fight one (either High Bias or High Variance)

can end up worsening the other. This is essentially how the Bias Variance

trade-off is encountered in practice