

# chapter-2-4

January 22, 2022

Run in Google Colab

## 1 Introduction to Machine Learning

### 1.1 Bias and Variance

Suppose there is a relationship between an independent variable  $x$  and a dependent variable  $y$ :

$$y = f(x) + \epsilon \quad (1)$$

Where  $\epsilon$  is an error term with mean zero and variance  $\sigma^2$ . The error term captures either genuine randomness in the data or noise due to measurement error.

Suppose we find a deterministic model for this relationship:

$$y = \hat{f}(x) \quad (2)$$

Now it comes a new data point  $x'$  not in the training set and we want to predict the corresponding  $y'$ . The error we will observe in our model at point  $x'$  is going to be

$$\hat{f}(x') - f(x') - \epsilon \quad (3)$$

There are two different sources of error in this equation. The first one is included in the factor  $\epsilon$ , the second one, more interesting, is due to what is in our training set. A robust model should give us the same prediction whatever data we used for training our model. Let's look at the average error:

$$E[\hat{f}(x')] - f(x') \quad (4)$$

where the expectation is taken over random samples of training data (having the same distribution as the training data).

This is the definition of the **bias**

$$\text{Bias}[\hat{f}(x')] = E[\hat{f}(x')] - f(x') \quad (5)$$

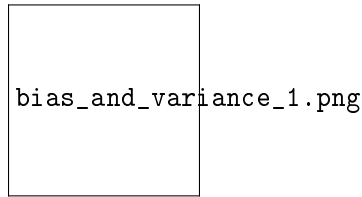
We can also look at the mean square error

$$E \left[ \left( \hat{f}(x') - f(x') - \epsilon \right)^2 \right] = \left[ \text{Bias} \left( \hat{f}(x') \right) \right]^2 + \text{Var} \left[ \hat{f}(x') \right] + \sigma^2 \quad (6)$$

Where we remember that  $\hat{f}(x')$  and  $\epsilon$  are independent.

This show us that there are two important quantities, the **bias** and the **variance** that will affect our results and that we can control to some extent.

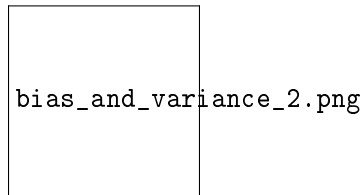
**FIGURE 1.1 - A good model should have low bias and low variance**



**Bias is how far away the trained model is from the correct result on average.** Where *on average* means over many goes at training the model using different data. And **Variance is a measure of the magnitude of that error.**

Unfortunately, we often find that there is a trade-off between bias and variance. As one is reduced, the other is increased. This is the matter of over- and under-fitting.

**Overfitting is when we train our algorithm too well on training data, perhaps having too many parameters for fitting.**



## 1.2 References

**S. Raschka and V. Mirjalili**, *"Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2"*, 3rd Edition. Packt Publishing Ltd, 2019.

**A. Géron**, *"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow"*, 2nd Edition. O'Reilly Media, 2019

[Scikit-Learn web site](#)