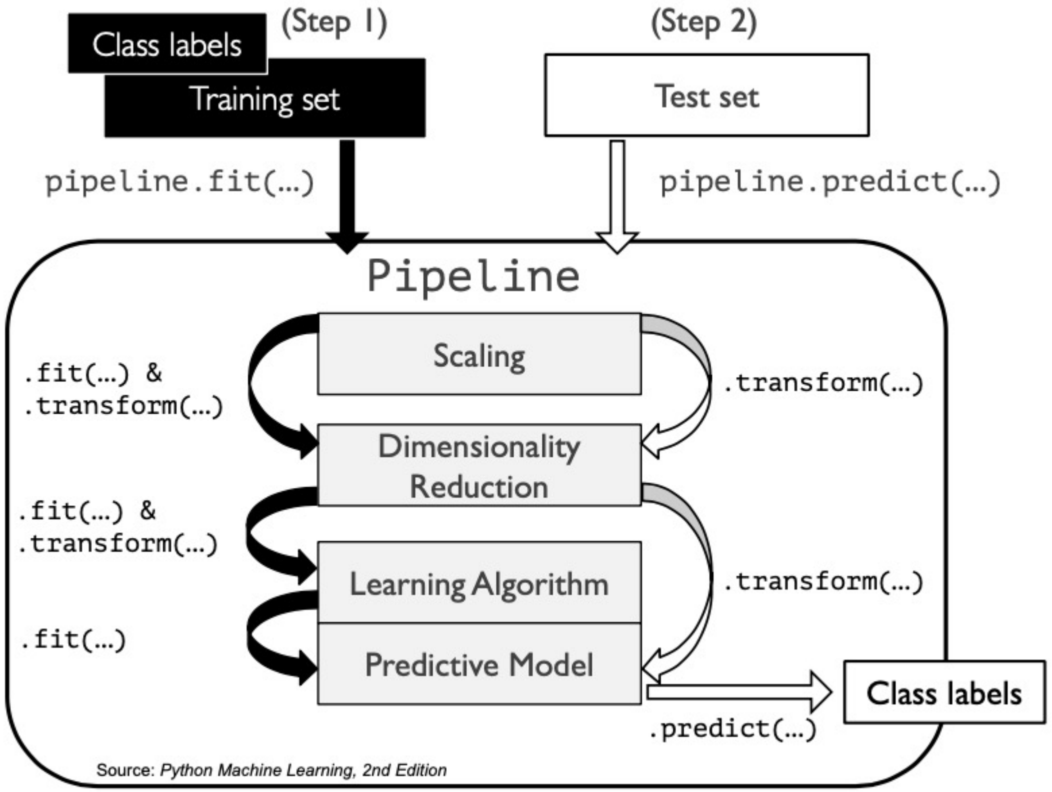
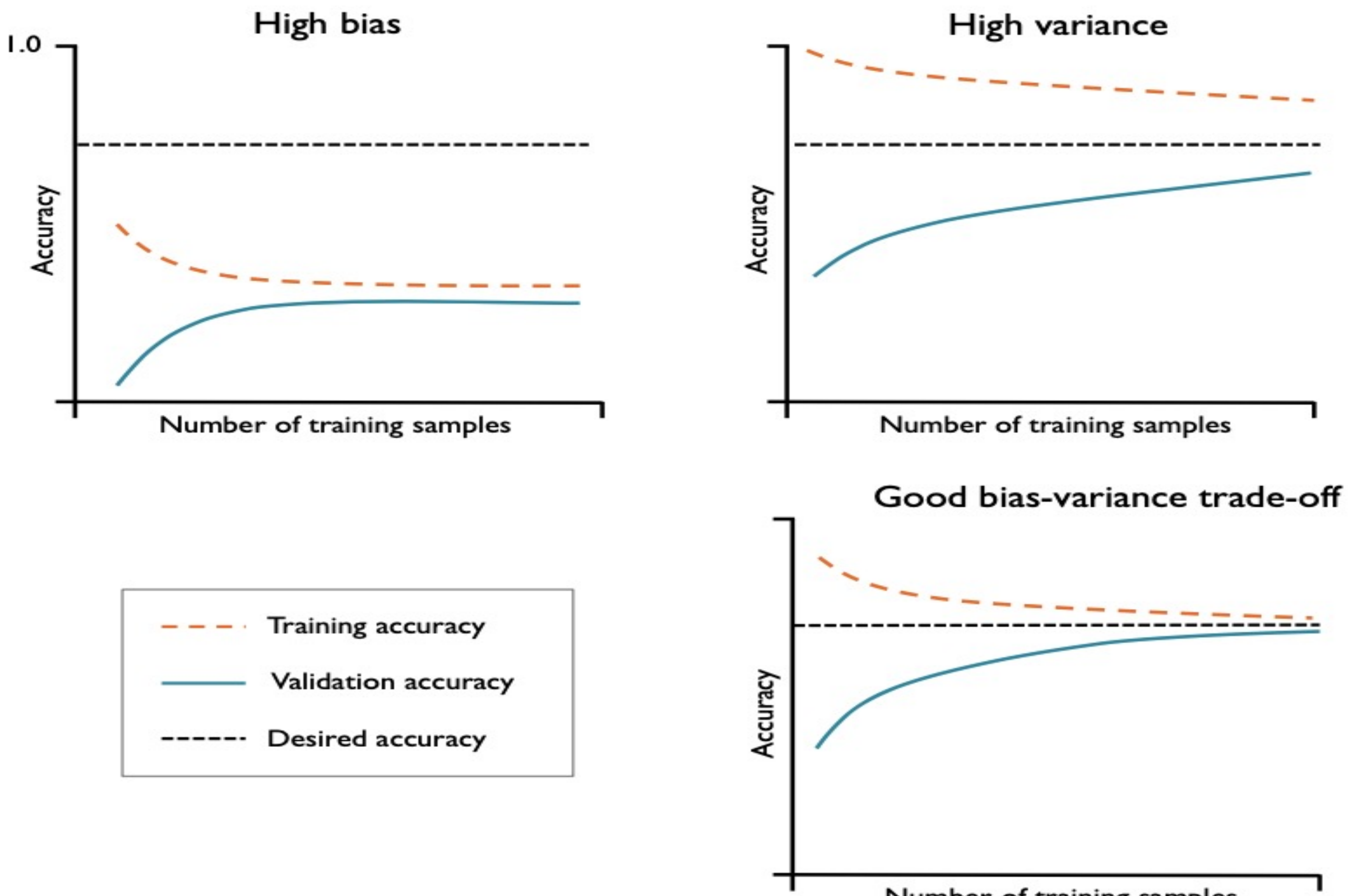
Model Evaluation

Model Evaluation



Bias – Variance Trade-off



Source: *Python Machine Learning, 2nd Edition*

2x2 Confusion Matrix

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

$$ERR = \frac{FP + FN}{FP + FN + TP + TN} = 1 - ACC \tag{1}$$

$$ACC = \frac{TP + TN}{FP + FN + TP + TN} = 1 - ERR \tag{2}$$

Confusion Matrix for Multi-Class Settings

Predicted Labels

True Labels

	Class 0	Class 1	Class 2
Class 0	T(0,0)		
Class 1		T(1,1)	
Class 2			T(2,2)

Confusions matrices are traditionally for binary class problems but we can easily generalize it to multi-class settings

False Positive Rate and False Negative Rate

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} \tag{3}$$

$$TPR = \frac{TP}{P} = \frac{TP}{FN + TP} \tag{4}$$

Precision, Recall, and F1 Score

$$PRE = \frac{TP}{TP + FP} \tag{5}$$

$$REC = TPR = \frac{TP}{P} = \frac{TP}{FN + TP} \tag{6}$$

$$F_1 = 2 \cdot \frac{PRE \cdot REC}{PRE + REC} \tag{7}$$

Sensitivity and Specificity

$$SEN = TPR = REC = \frac{TP}{P} = \frac{TP}{FN + TP} \tag{8}$$

$$SPC = TNR = \frac{TN}{N} = \frac{TN}{FP + TN} \tag{9}$$

Sensitivity measures the recovery rate of the Positives and complimentary, the Specificity measures the recovery rate of the Negatives.

Balanced Accuracy / Average Per-Class Accuracy

Predicted Labels

True Labels

	Class 0	Class 1	Class 2
Class 0	T(0,0)		
Class 1		T(1,1)	
Class 2			T(2,2)

$$ACC = \frac{T}{n}$$

Predicted Labels

True Labels

	Class 0	Class 1	Class 2
Class 0	3	0	0
Class 1	7	50	12
Class 2	0	0	18

$$ACC = \frac{3 + 50 + 18}{90} \approx 0.79$$

$$APC\ ACC = \frac{83/90 + 71/90 + 78/90}{3} \approx 0.86$$

Balanced Accuracy / Average Per-Class Accuracy

Predicted Labels (columns) / True Labels (rows)

	Class 0	Neg Class
Class 0	3	0
Neg Class	7	80

Predicted Labels (columns) / True Labels (rows)

	Class 1	Neg Class
Class 1	50	19
Neg Class	0	21

Predicted Labels (columns) / True Labels (rows)

	Class 2	Neg Class
Class 2	18	0
Neg Class	12	60

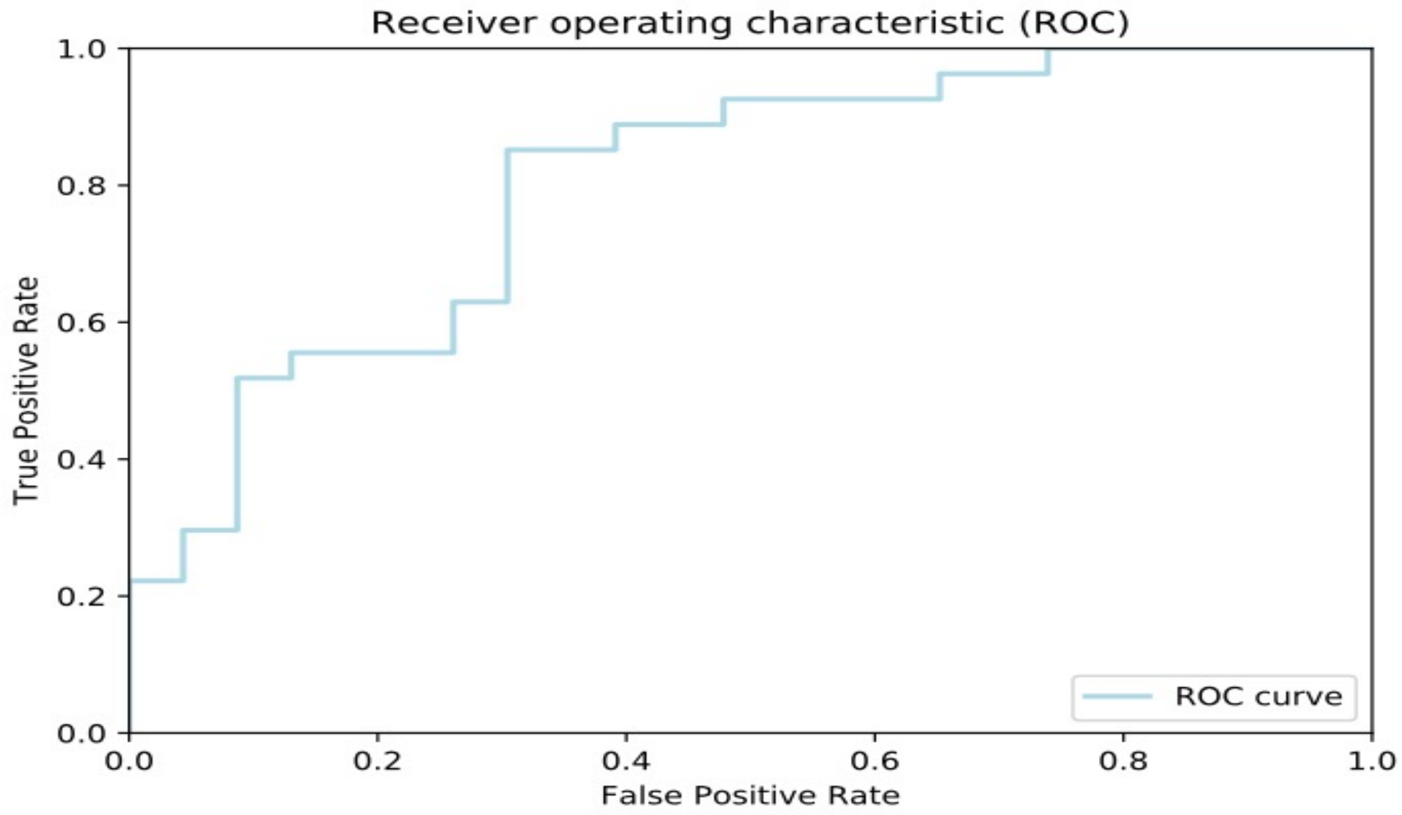
Predicted Labels (columns) / True Labels (rows)

	Class 0	Class 1	Class 2
Class 0	3	0	0
Class 1	7	50	12
Class 2	0	0	18

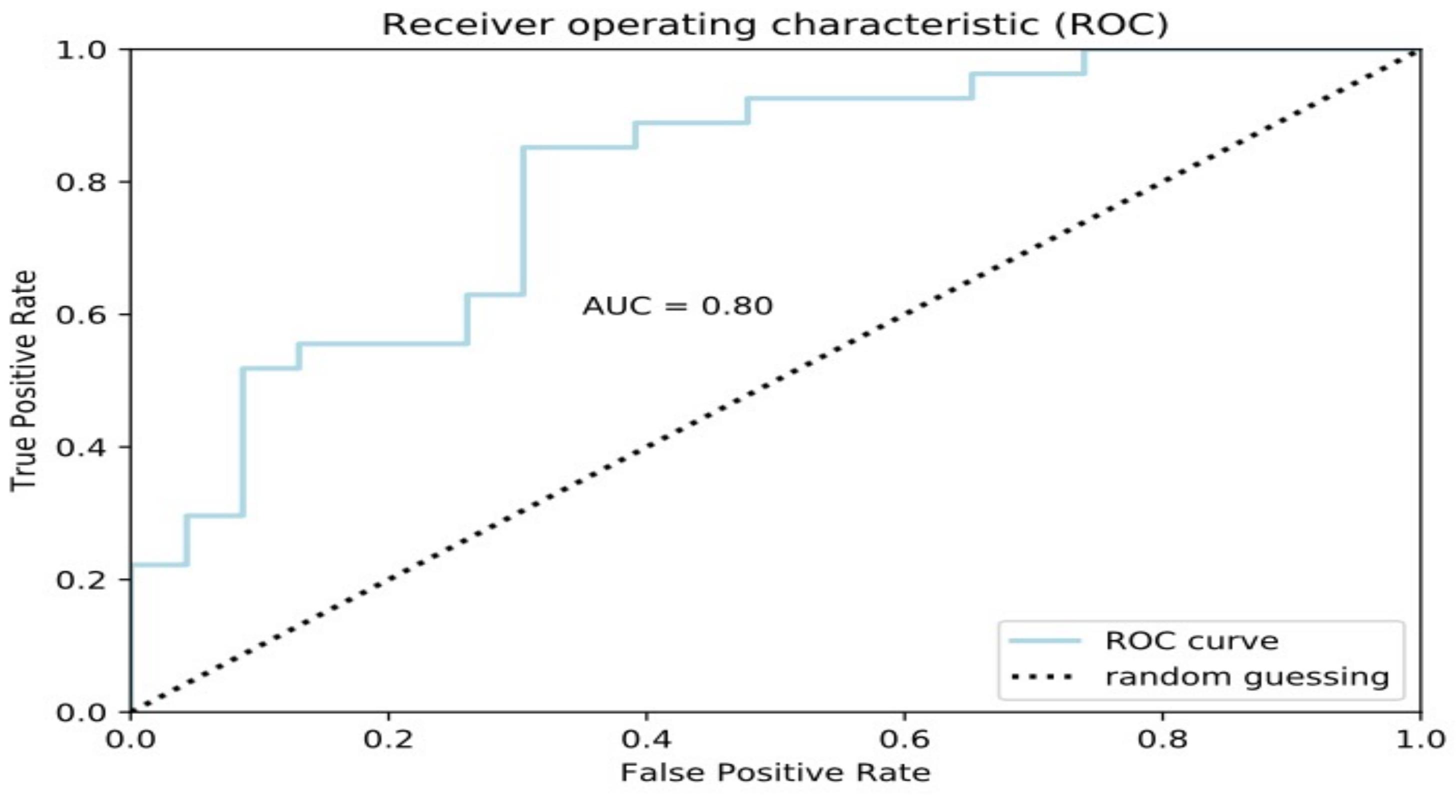
$$APC\ ACC = \frac{83/90 + 71/90 + 78/90}{3} \approx 0.86$$

Receiver Operating Characteristic curve (ROC curve)

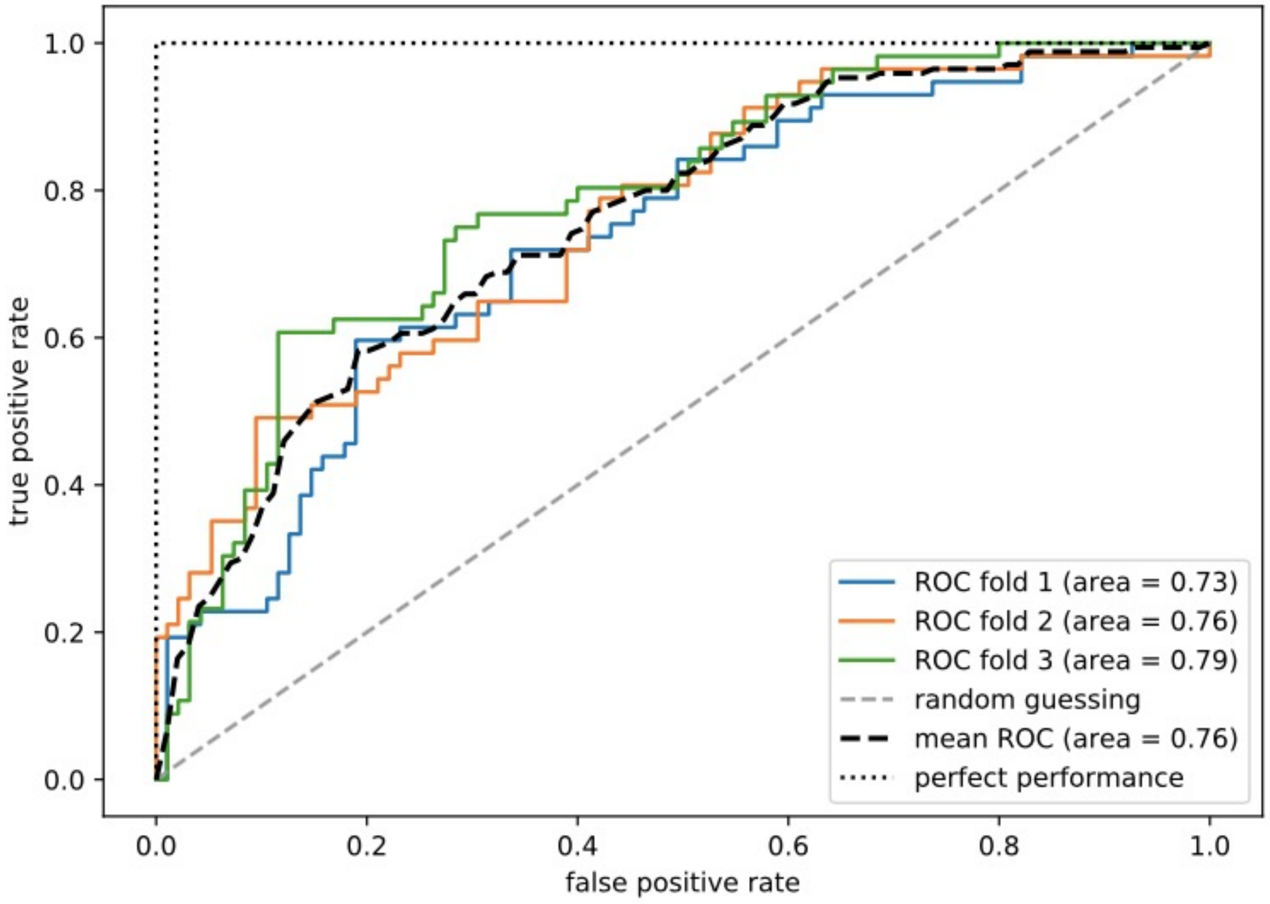
- Trade-off between True Positive Rate and False Positive Rate
- ROC can be plotted by changing the prediction threshold
- ROC term comes from "Radar Receiver Operators"
  (analysis of radar [**RA**dio **D**irection **A**nd **R**anging] images)



ROC Area Under the Curve (AUC)



ROC and k-Fold Cross-Validation



Macro and Micro Averaging

$$PRE_{micro} = \frac{TP_1 + \ldots + TP_c}{TP_1 + \ldots + TP_c + FP_1 + \ldots + FP_c}$$

$$PRE_{macro} = \frac{PRE_1 + \ldots + PRE_c}{c}$$

Micro-averaging is useful if we want to weight each instance or prediction equally, whereas macro-averaging weights all classes equally to evaluate the overall performance of a classifier with regard to the most frequent class labels.