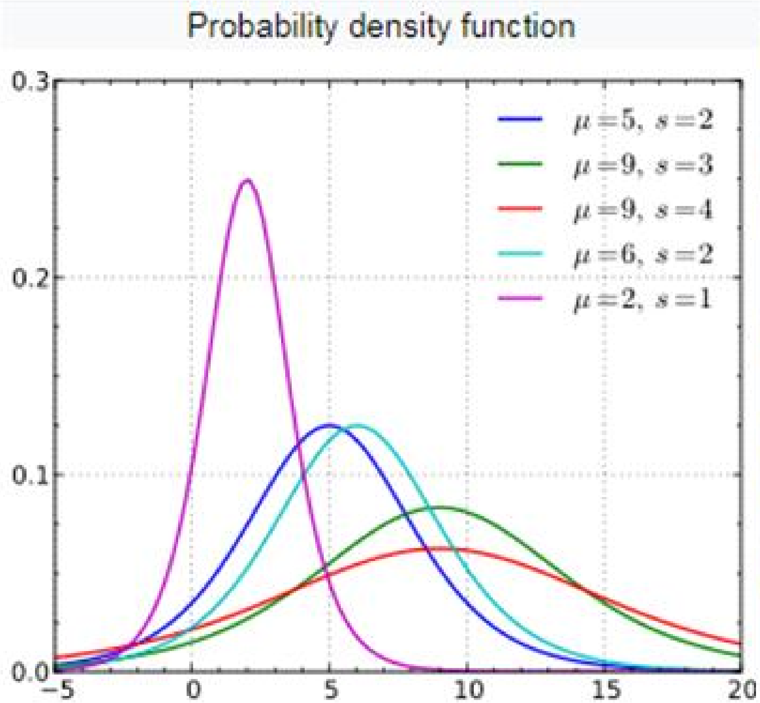
Logistic Regression + ROC Curves
(all figures from StatQuest)

Nazerfard, Ehsan
nazerfard@aut.ac.ir

Logistic Distribution



Parameters	μ, location (real) $s > 0$, scale (real)
Support	$x \in (-\infty, \infty)$
PDF	$\dfrac{e^{-(x-\mu/)s}}{s\left(1+e^{-(x-\mu)/s}\right)^2}$
CDF	$\dfrac{1}{1+e^{-(x-\mu)/s}}$
Mean	μ
Median	μ
Mode	μ
Variance	$\dfrac{s^2\pi^2}{3}$

Figure © Wikipedia

Logistic Function

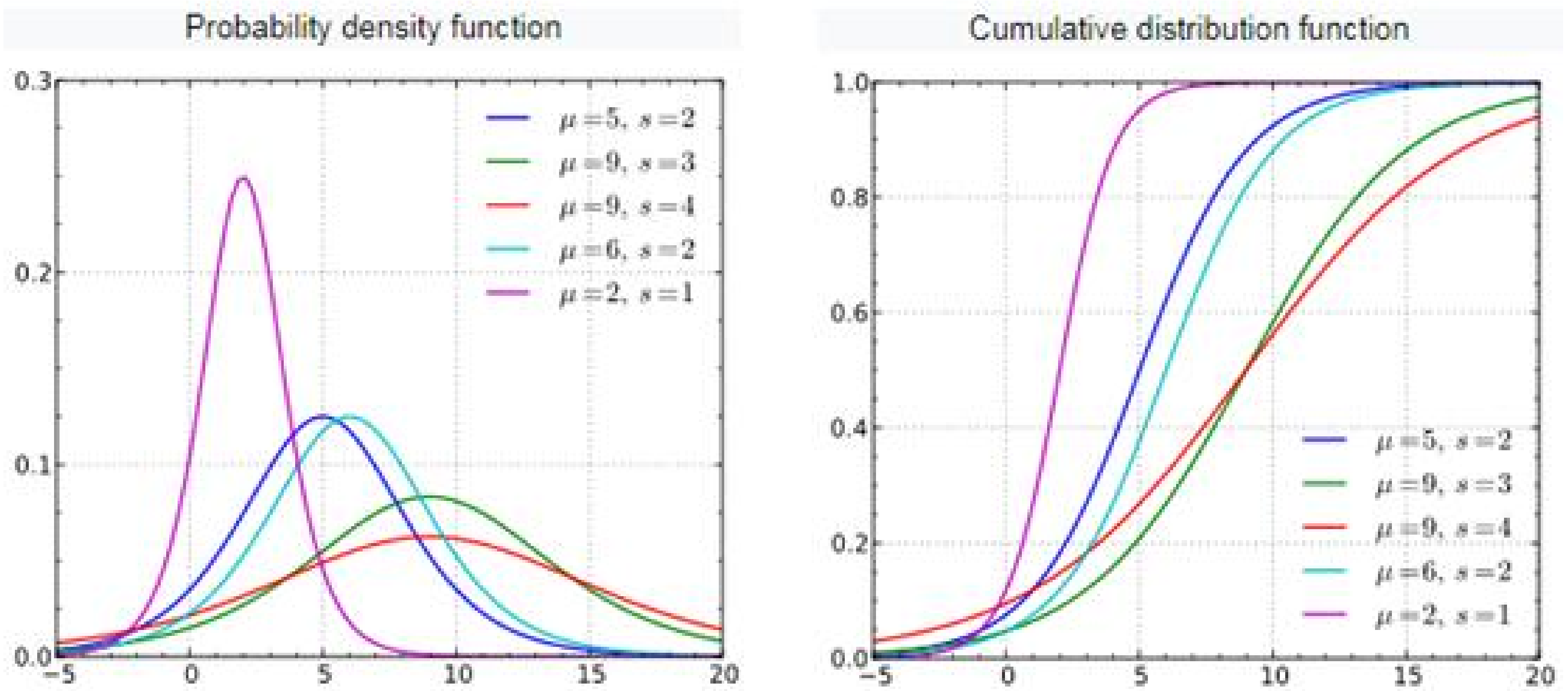
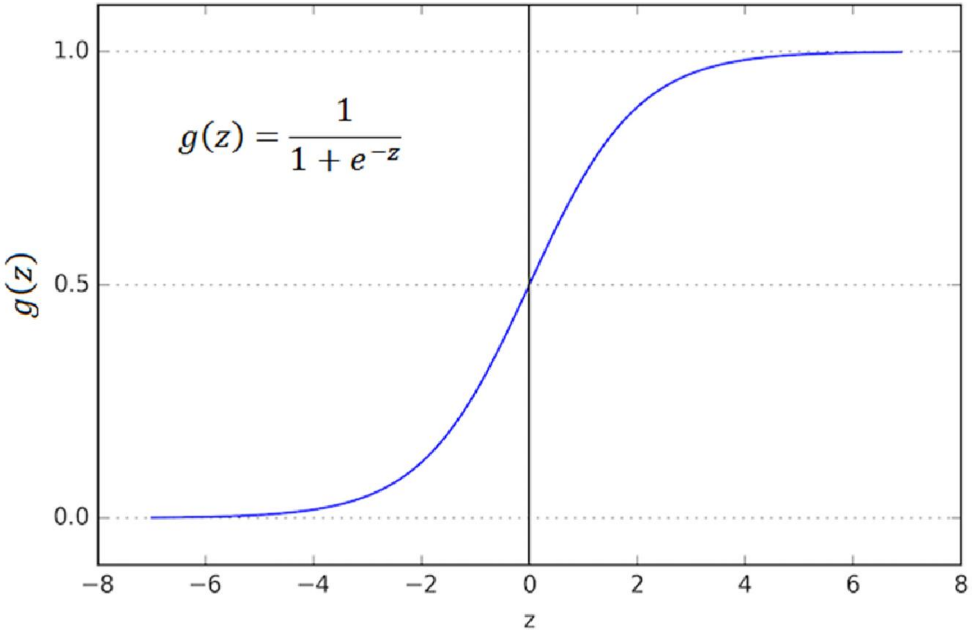


Figure © Wikipedia

Logistic Function (cont.)



$$\frac{dg(z)}{dz} = \mathrm{g(z)}[1 - \mathrm{g(z)}]$$

Idea

- Let's use the non-linear regression algorithm for a classification task.
  - Consider the Tumor Prognosis problem.

Sample Tumor Data

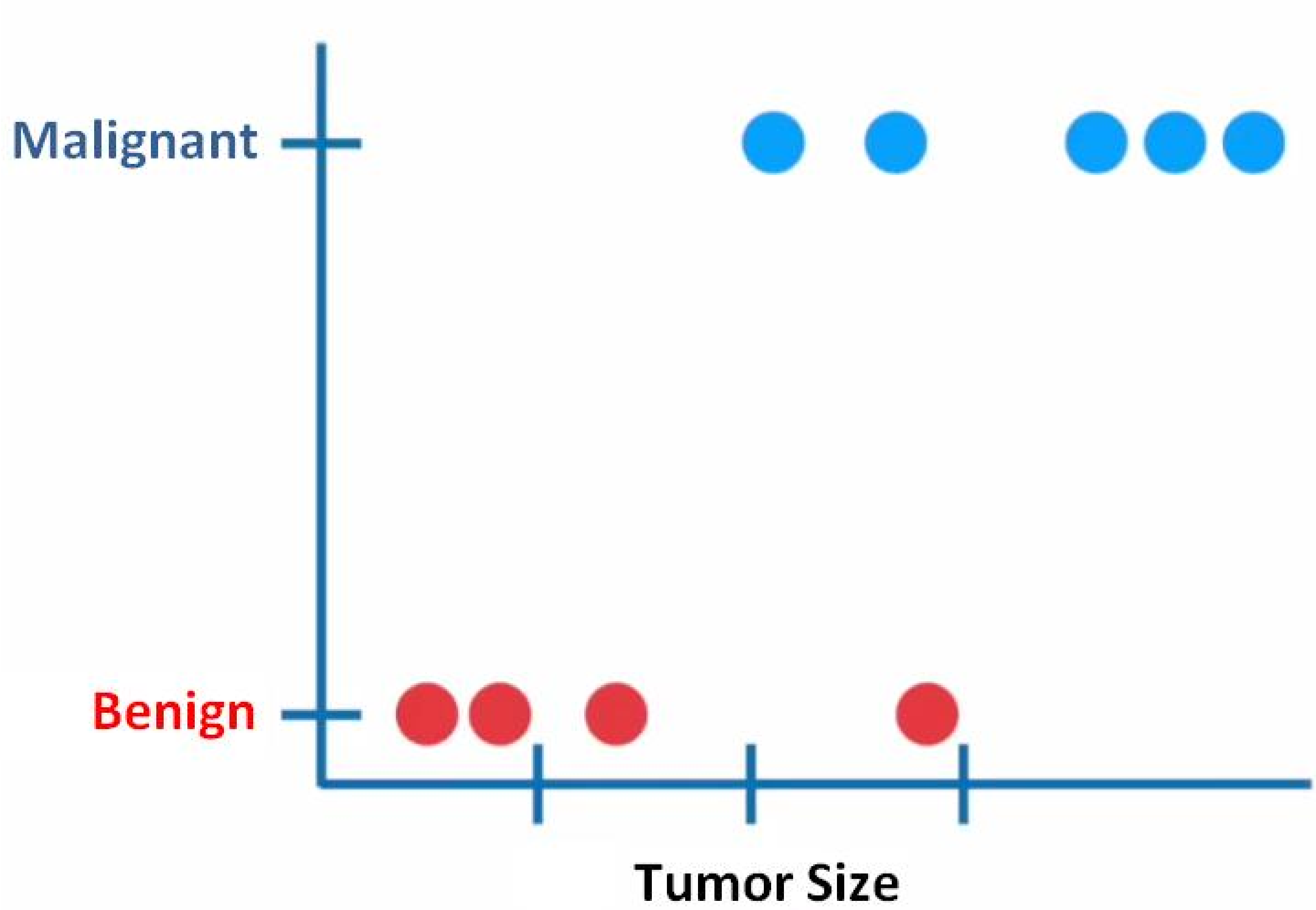
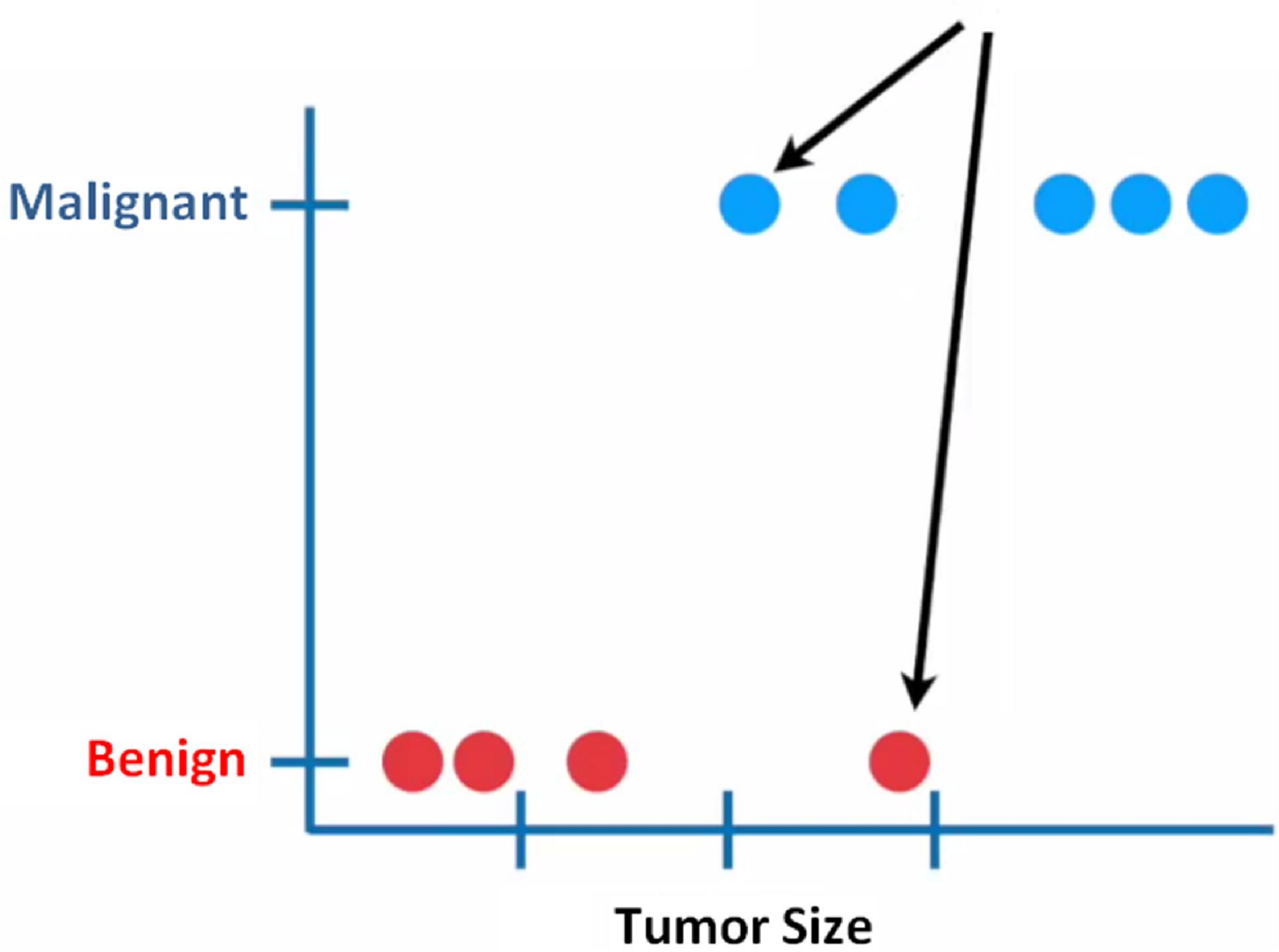
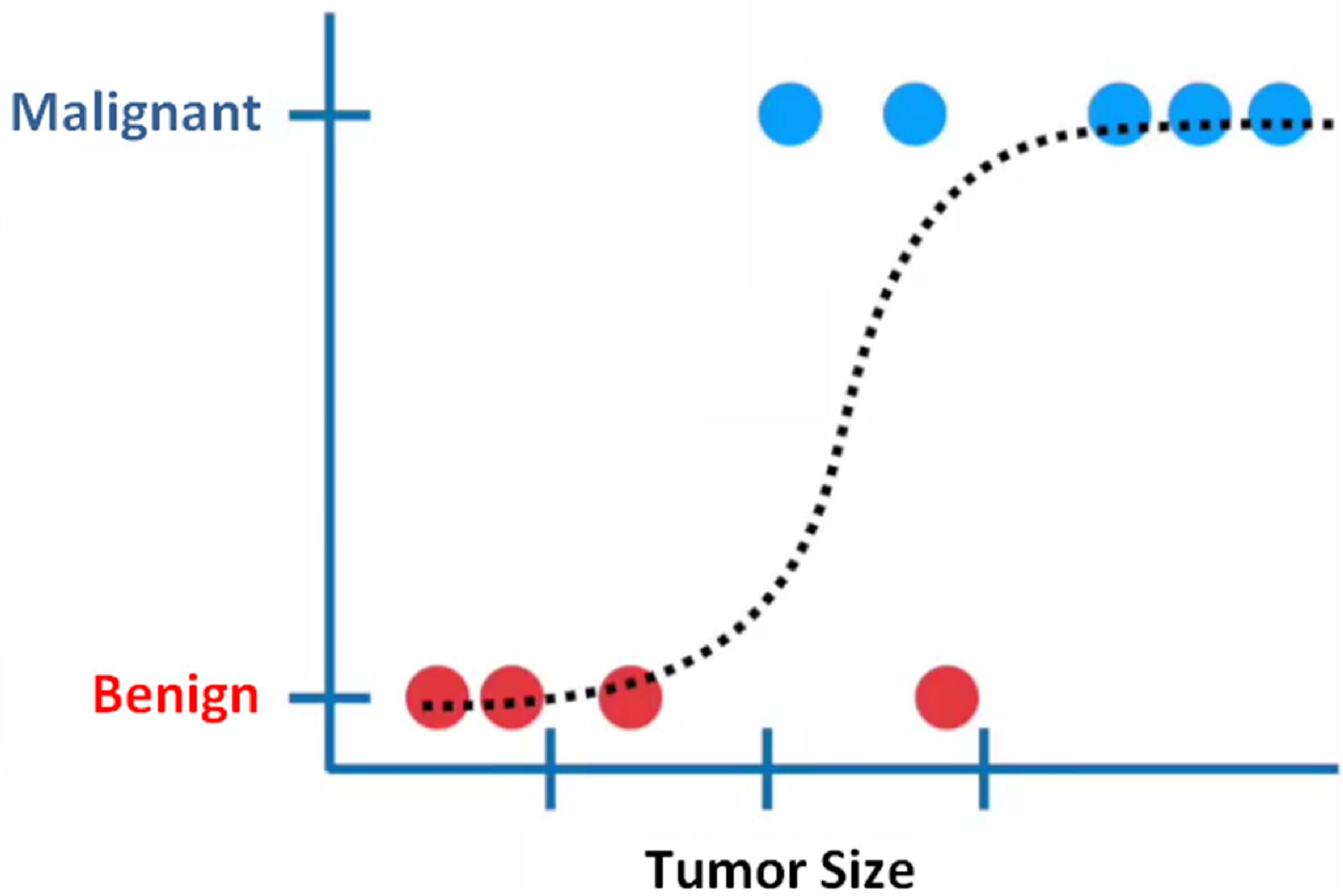


Figure © statQuest

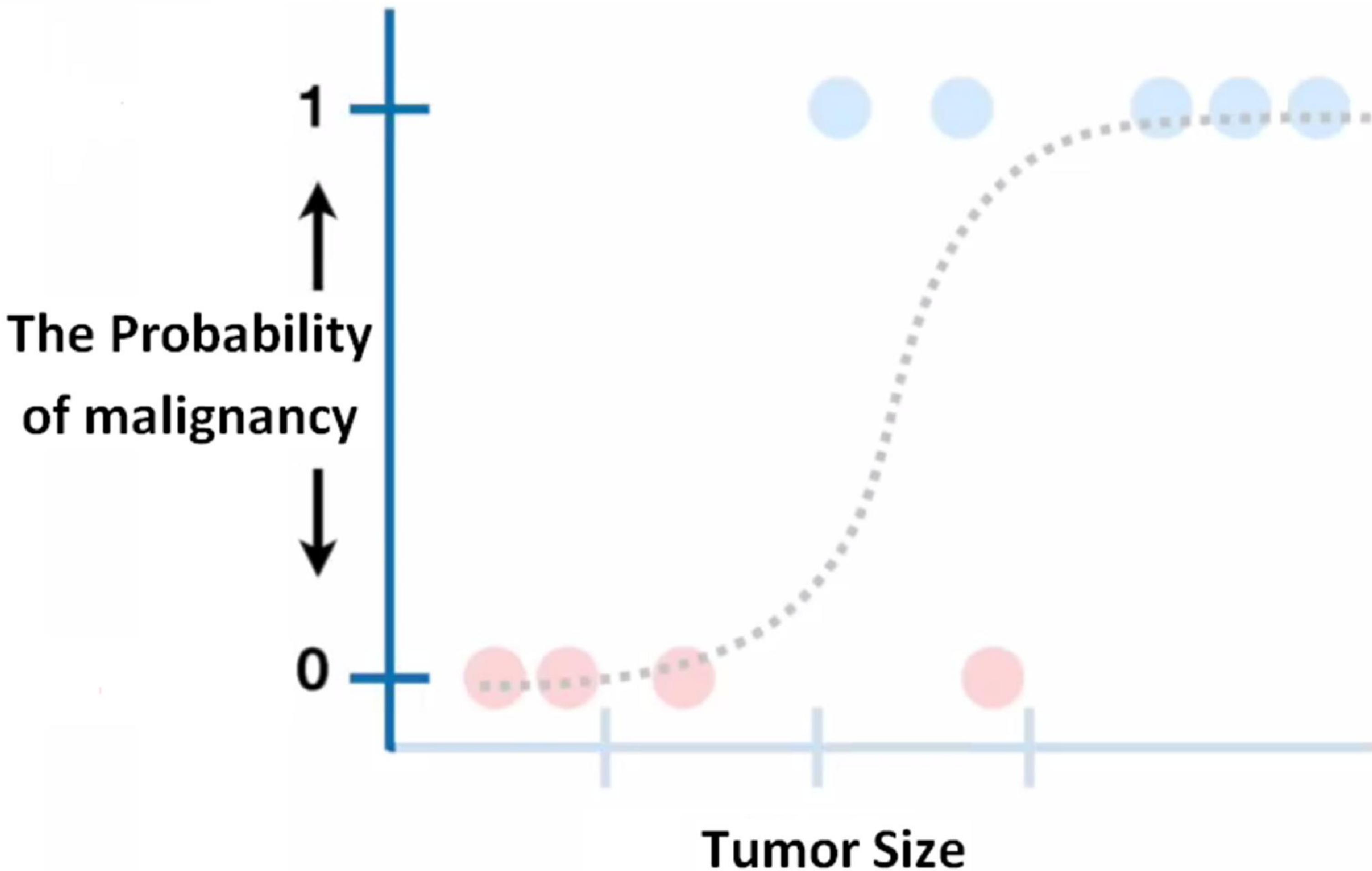
Sample Data



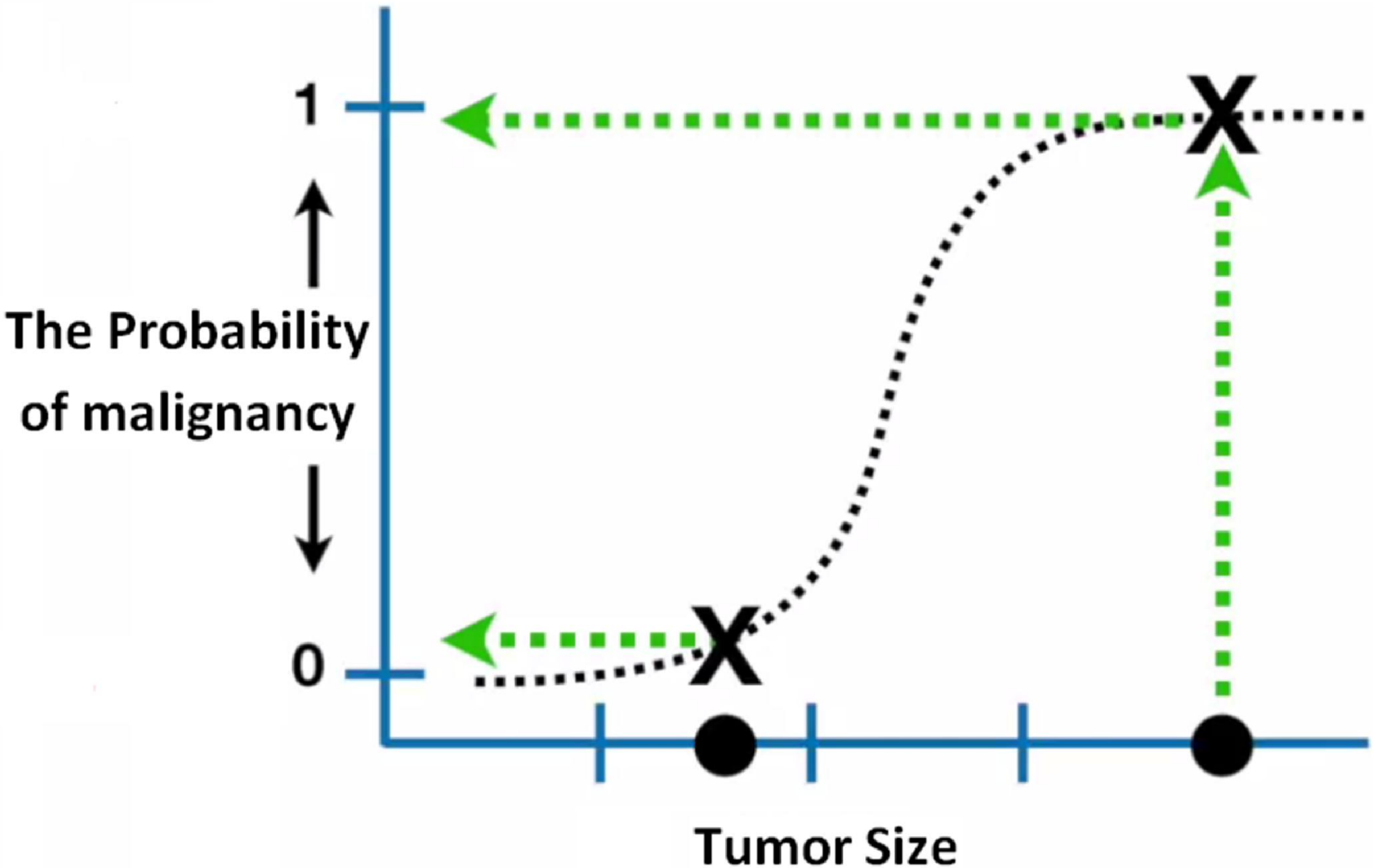
Fitting a Logistic Function



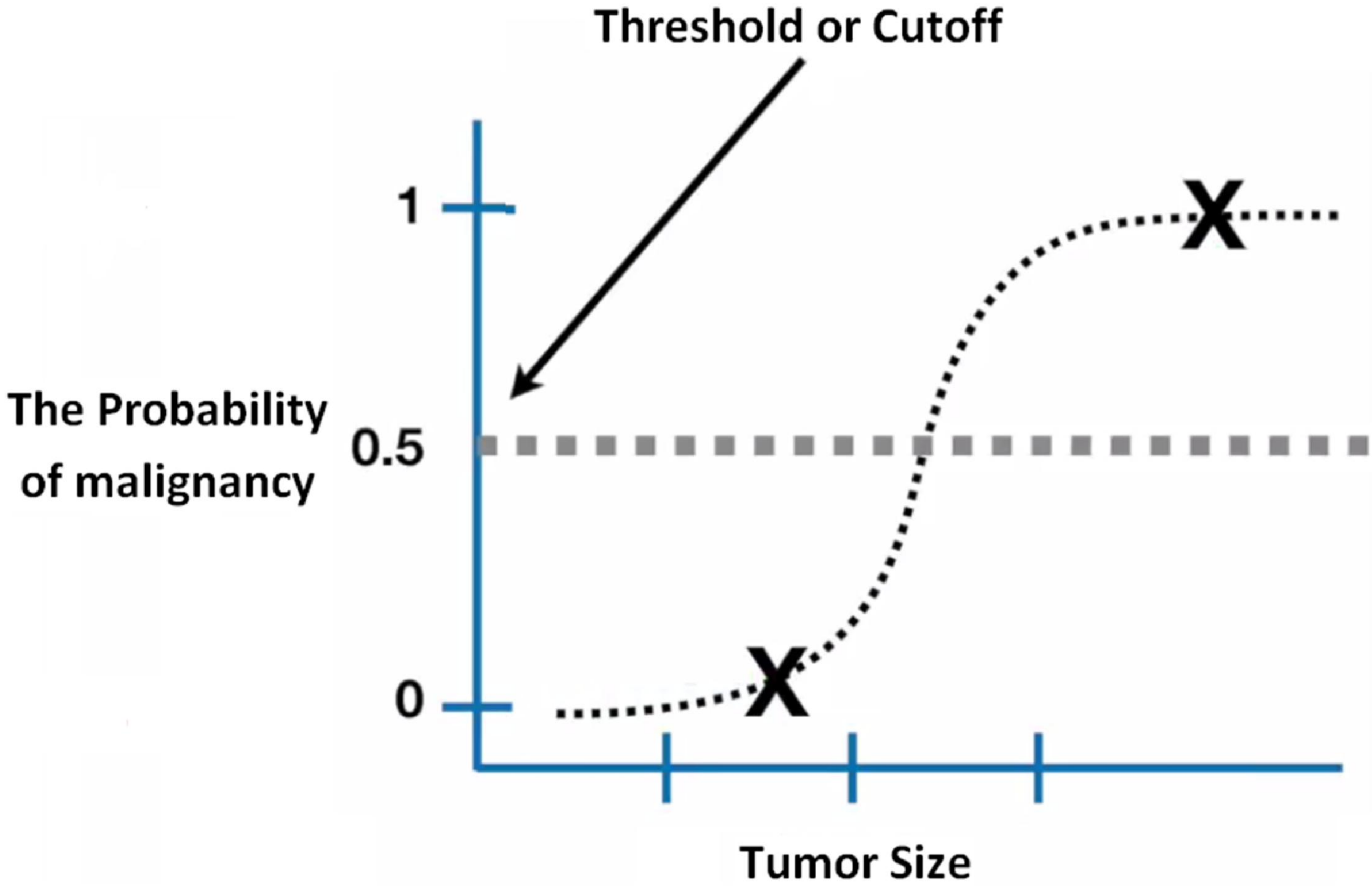
The Y-axis



Logistic Regression

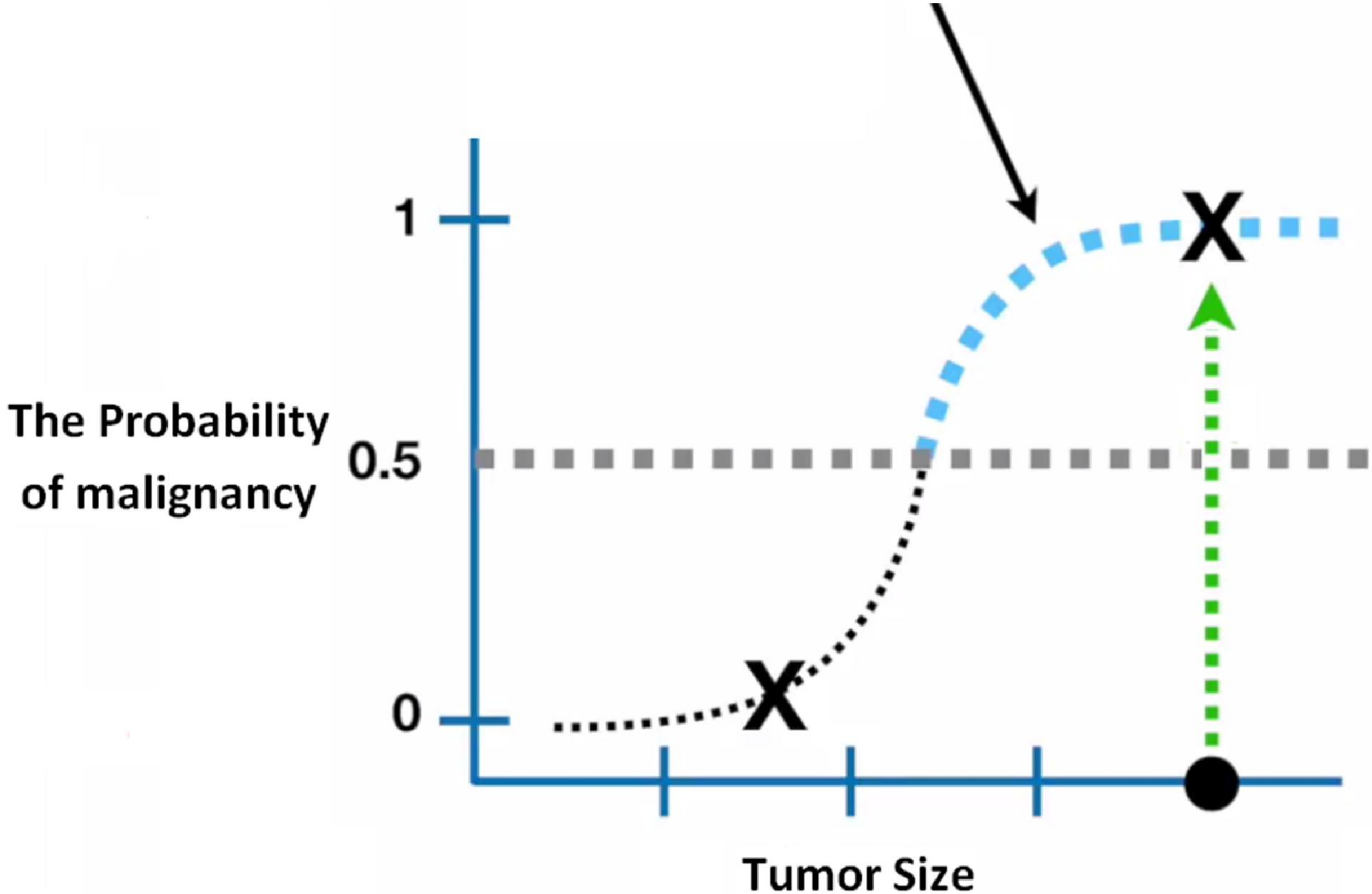


Logistic Regression for Classification



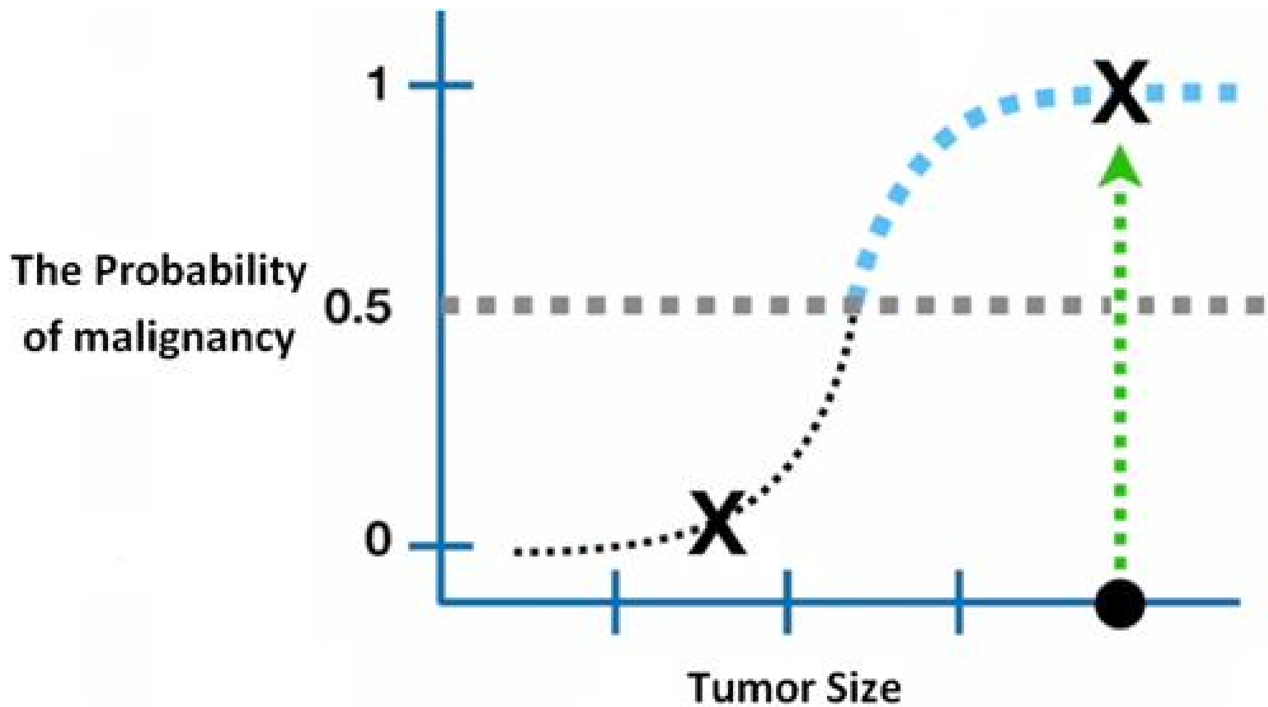
Logistic Regression for Classification (cont.)

- If $Probability(x) \geq 0.5$ → The Tumor is malignant
  - x is a tumor size



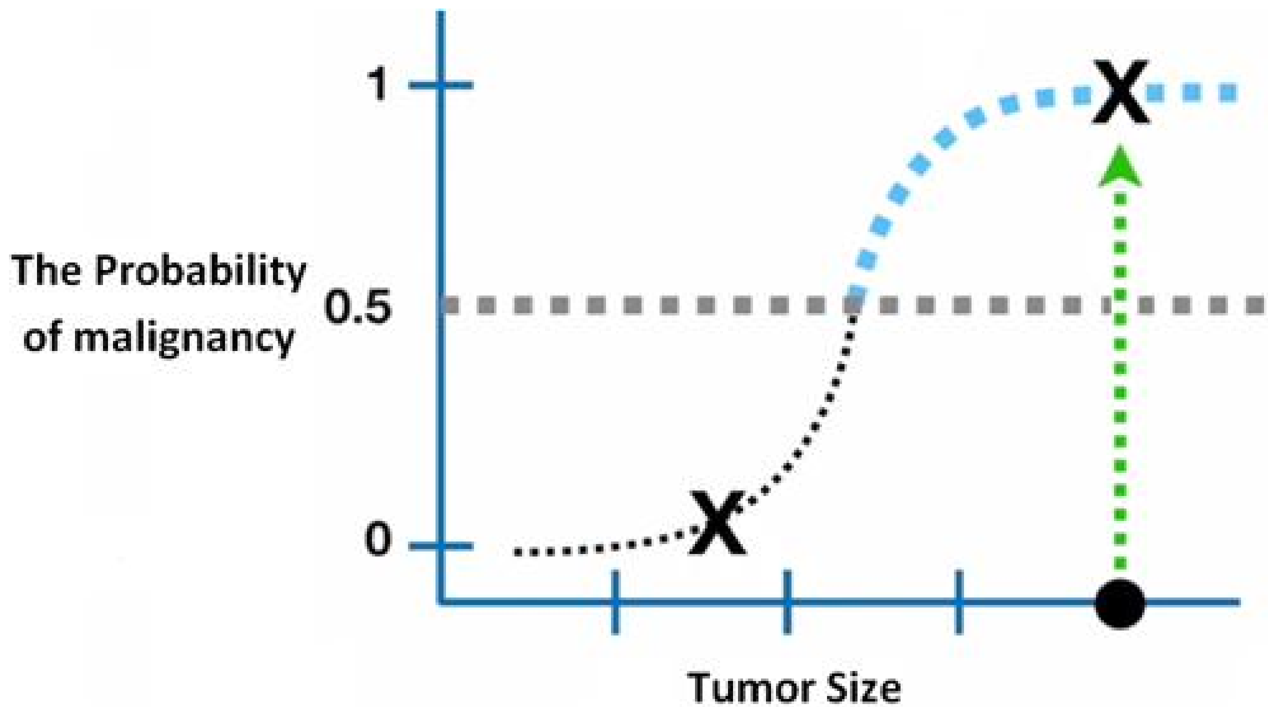
Logistic Regression for Classification (cont.)

- $p(Y = 1|X = x) = \frac{1}{1+\exp(-x)}$
  - x is a tumor size



Logistic Regression for Classification (cont.)

- $p(Y=1|X=x)=\frac{1}{1+\exp(-x)} \rightarrow p(Y=0|X=x)=1-\frac{1}{1+\exp(-x)}$
  - x is a tumor size



Logistic Regression for Classification (cont.)

- $p(Y=1|X=x)=\frac{1}{1+\exp(-x)} \rightarrow p(Y=0|X=x)=\frac{\exp(-x)}{1+\exp(-x)}$
  - x is a tumor size

- The precise form of a Gaussian naïve Bayes is obtained!
- As the # of training examples grow toward infinity, Gaussian naïve Bayes and Logistic Regression converge toward identical classifiers.

Logistic Regression for Classification (cont.)

- $p(Y = 1|X = x) = \frac{1}{1+\exp(-x)} \rightarrow p(Y = 0|X = x) = \frac{\exp(-x)}{1+\exp(-x)}$
  - x is a tumor size

- The precise form of a Gaussian naïve Bayes is obtained!
- As the # of training examples grow toward infinity, Gaussian naïve Bayes and Logistic Regression converge toward identical classifiers.

- Logistic regression directly estimates the parameters of $P(Y|X)$, whereas naïve Bayes directly estimates parameters for $P(Y)$ and $P(X|Y)$.
- The former is called a discriminative classifier, and the latter a generative classifier.

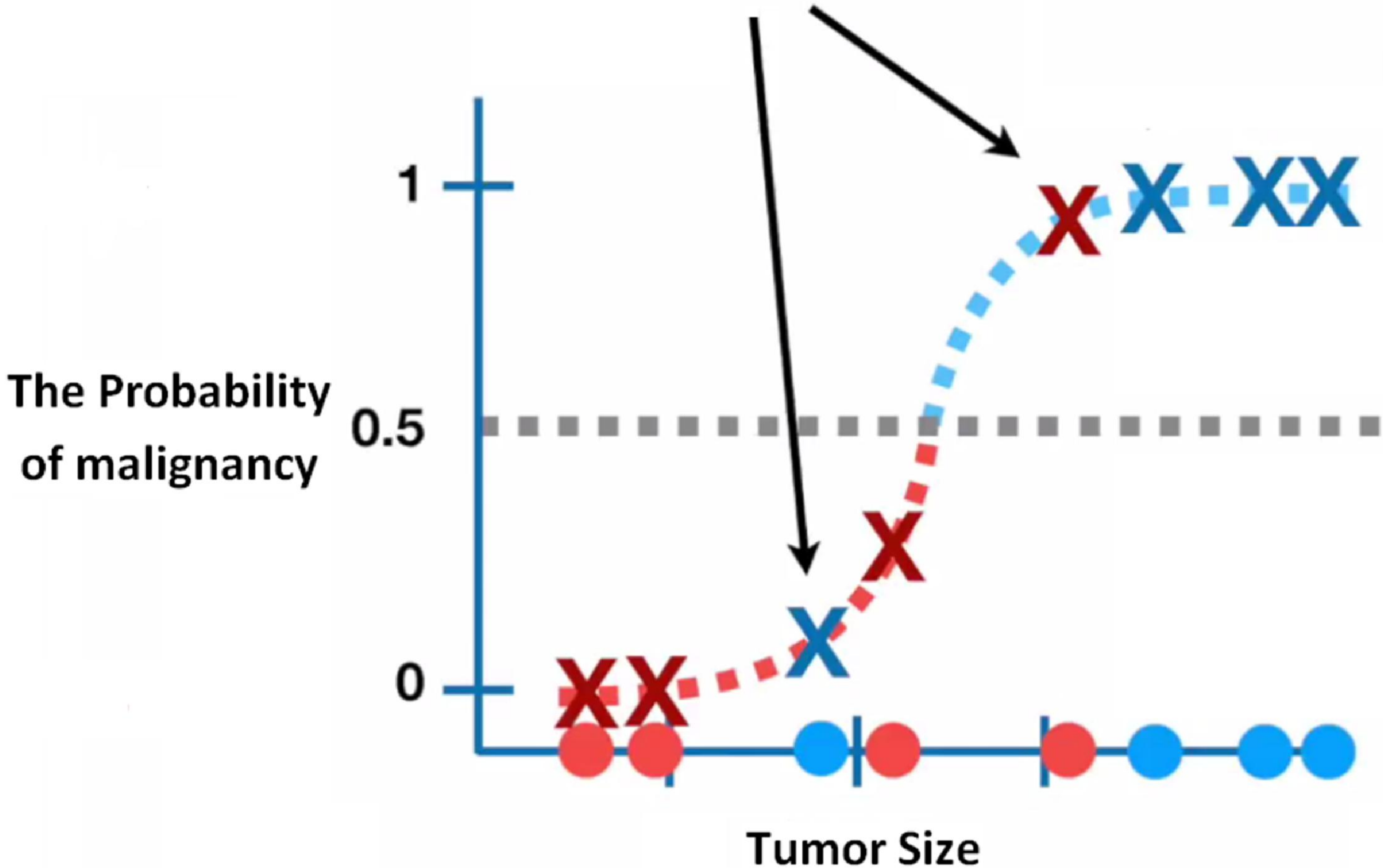
Further Reading

- Andrew Ng., Michael Jordan, On Discriminative vs. Generative classifiers: A comparison of logistic regression and naïve Bayes, NIPS 2001.

Further Reading (cont.)

- Generative Adversarial Networks (GANs)
- I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, Advances in neural information processing systems 27, 2672-2680.

Evaluation: False Predictions

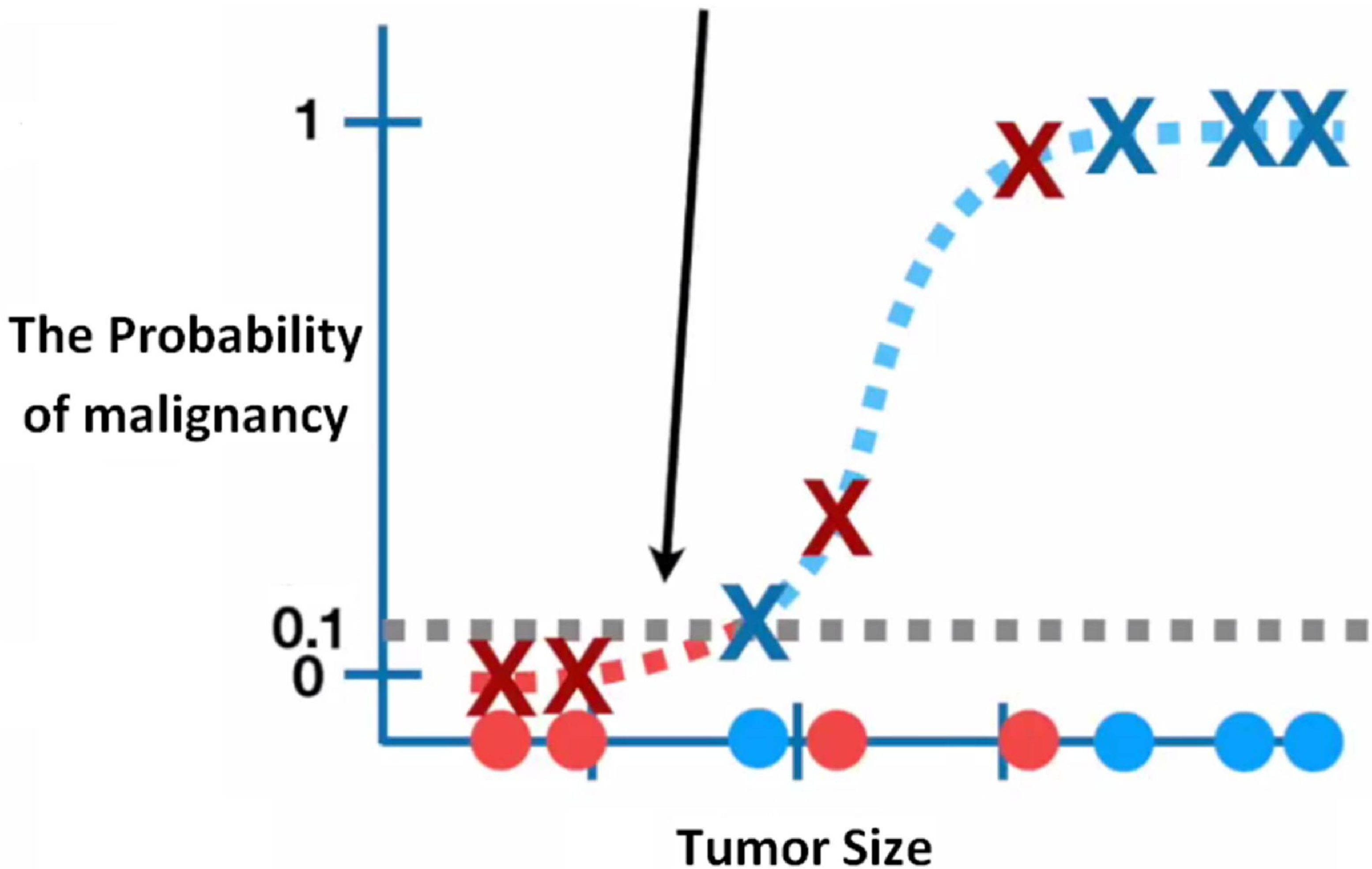


Confusion Matrix

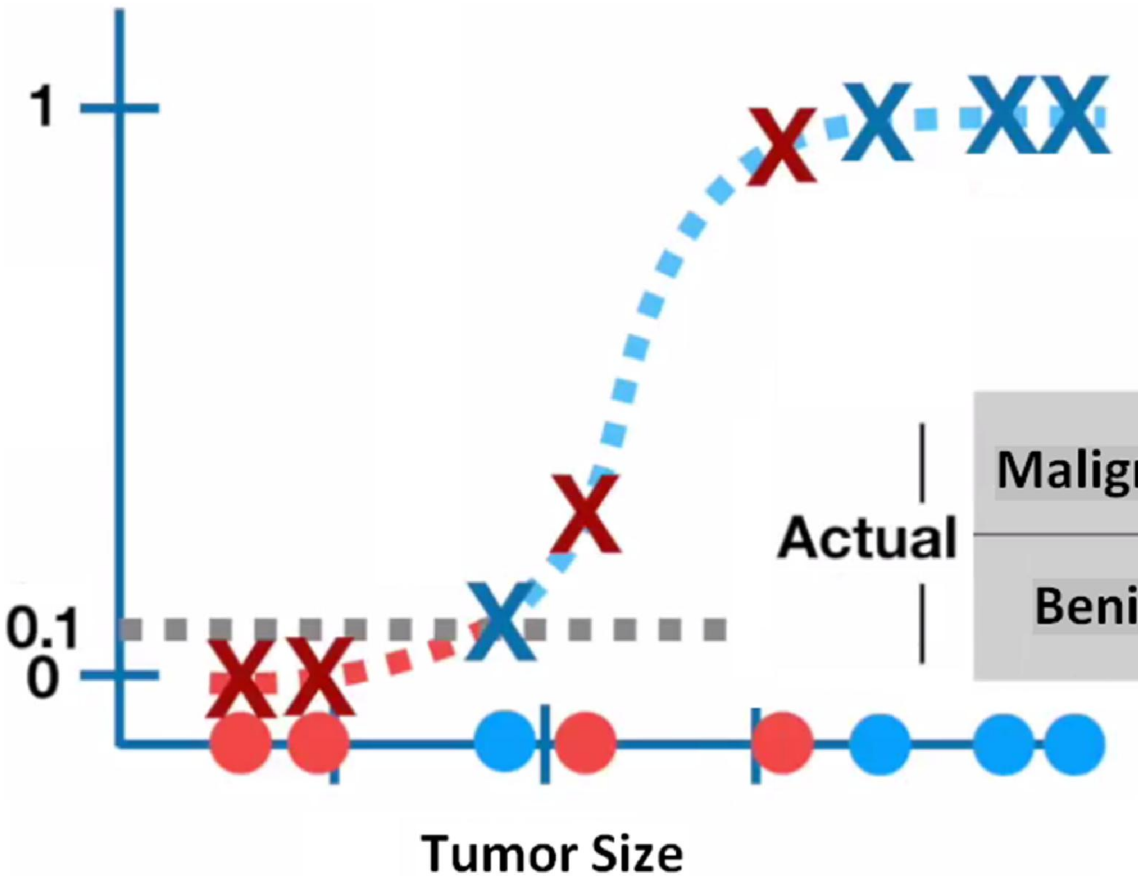
- Confusions corresponding the Threshold = 0.5

		Predicted	
		Malignant	Benign
Actual	Malignant	3	1
	Benign	1	3

Threshold is set to 0.1

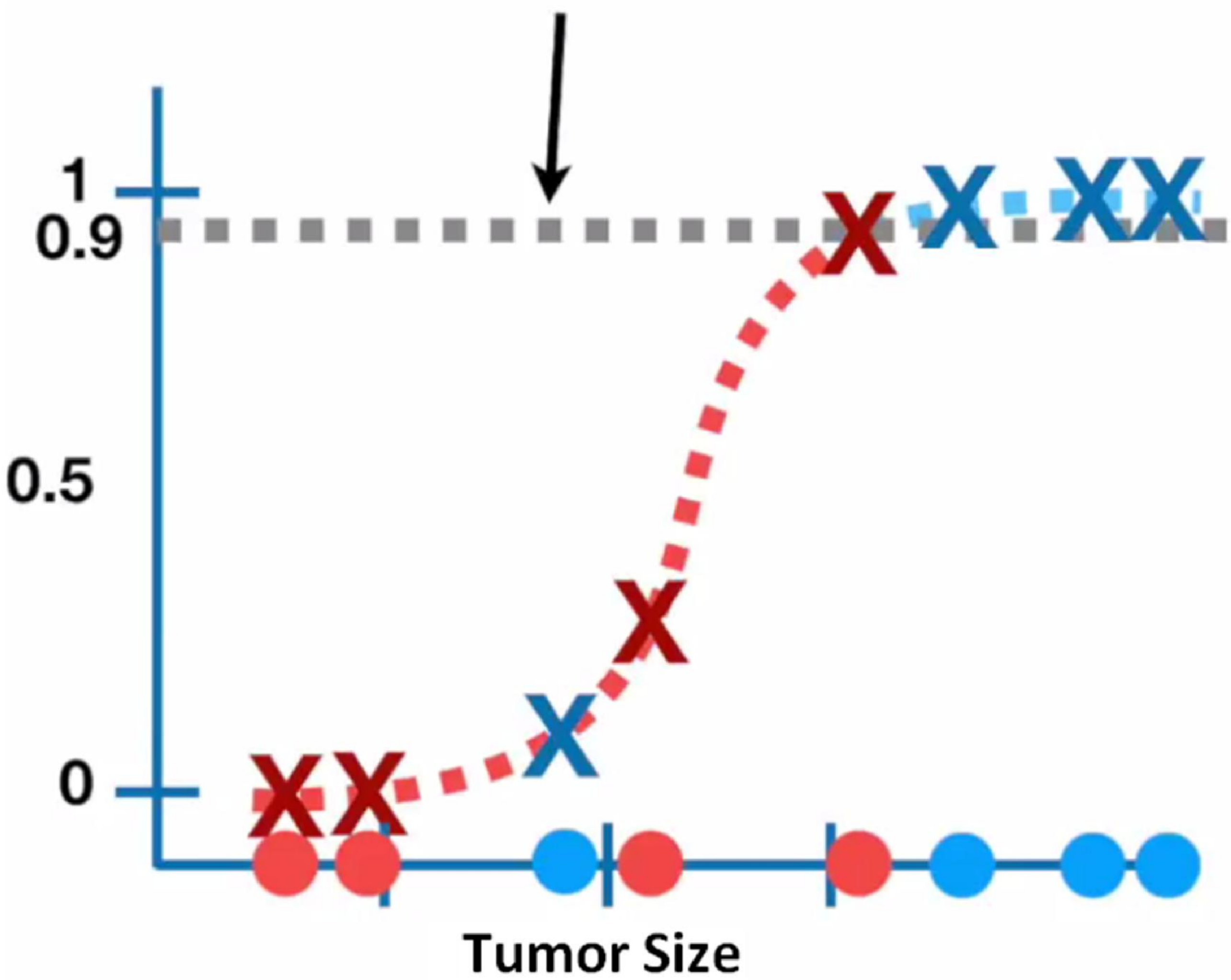


Corresponding Confusion Matrix

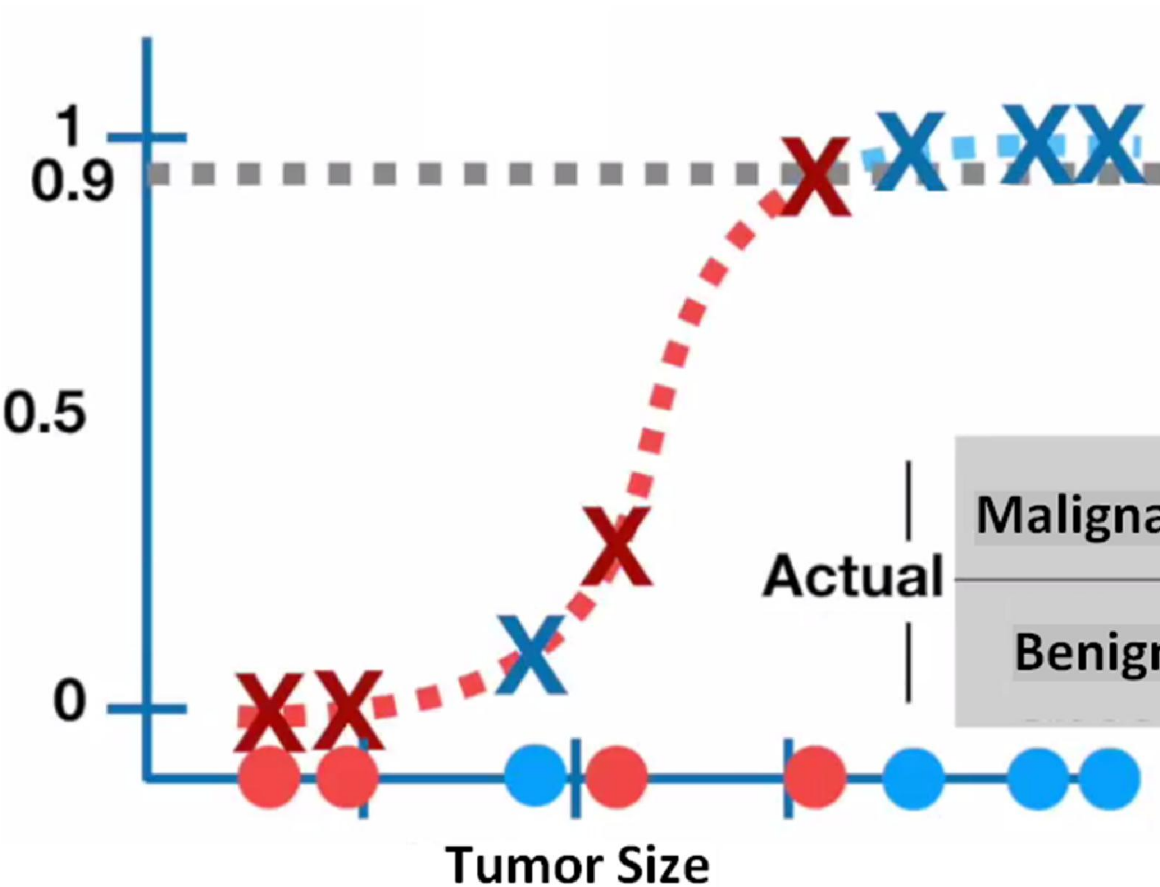


		Predicted	
		Malignant	**Benign**
Actual	**Malignant**	4	0
	Benign	2	2

Threshold is set to 0.9



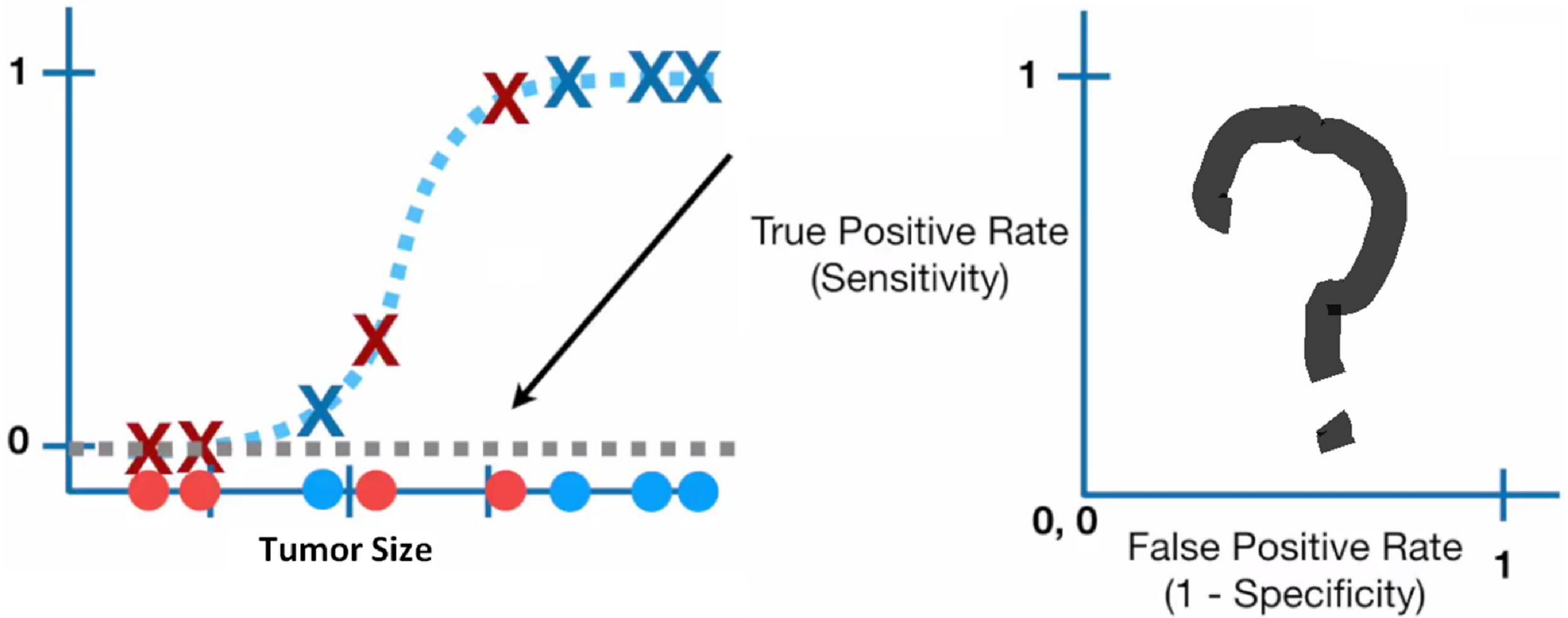
Corresponding Confusion Matrix



		Predicted	
		Malignant	Benign
Actual	Malignant	3	1
	Benign	0	4

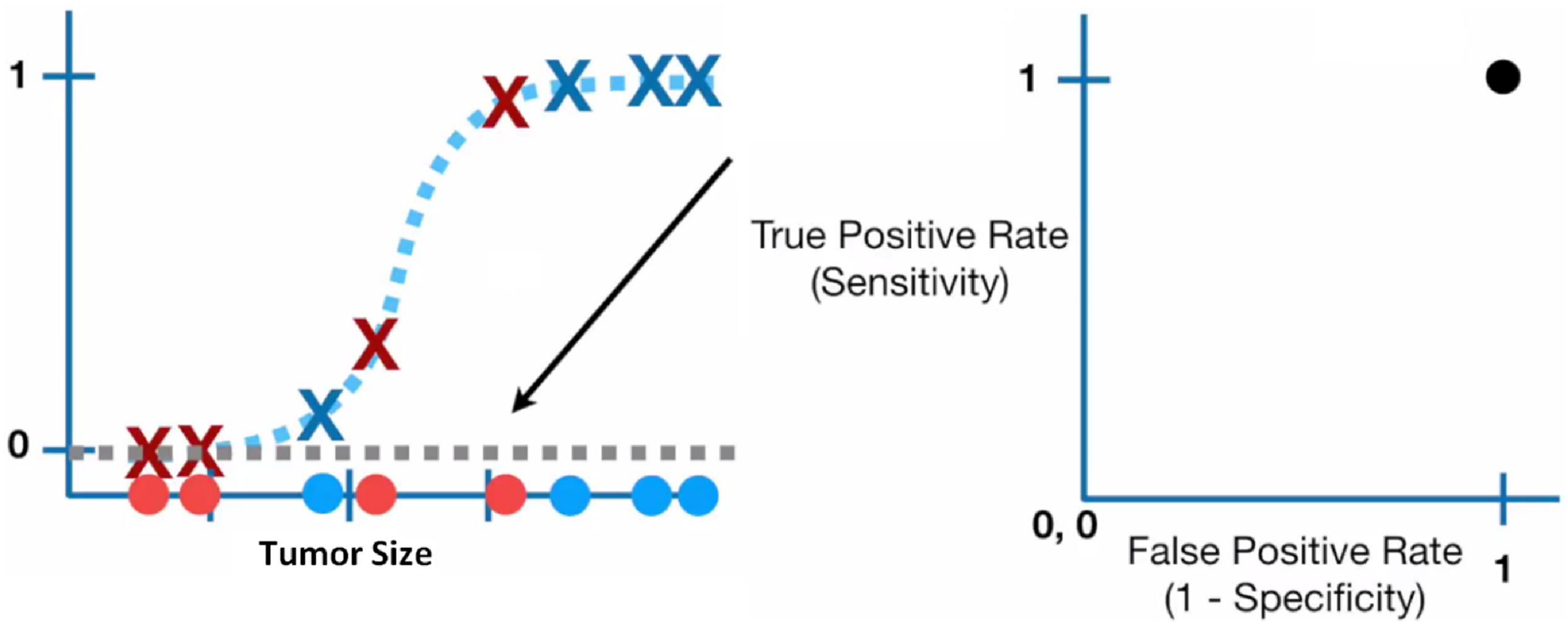
ROC Curves

- Receiver operator characteristic curves (ROC curves) provide a way to summarize all confusion matrices correspond to different thresholds.
- Starting with the threshold that classifies all samples as malignant.

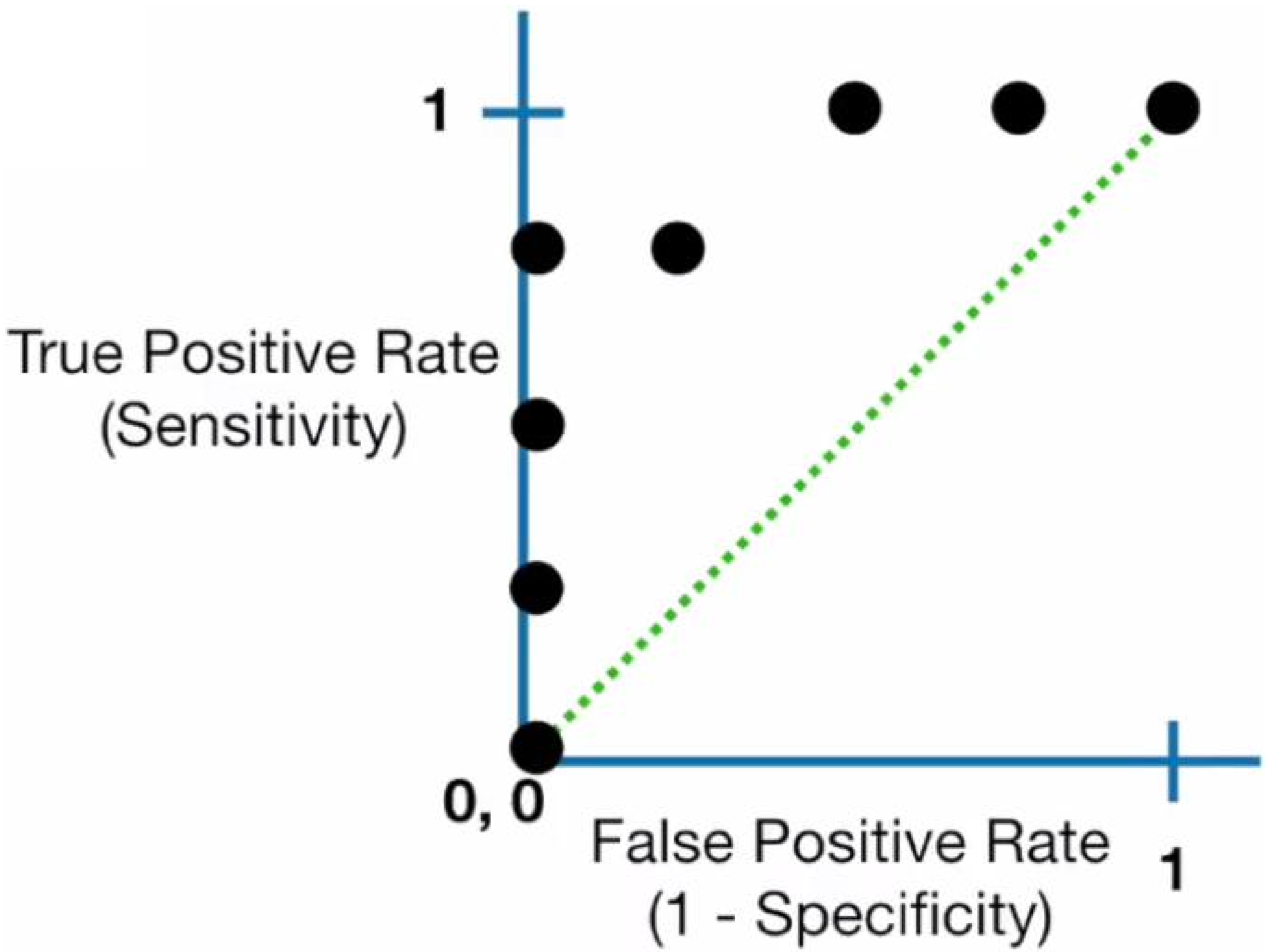


ROC Curves (cont.)

- Corresponding point to threshold zero.

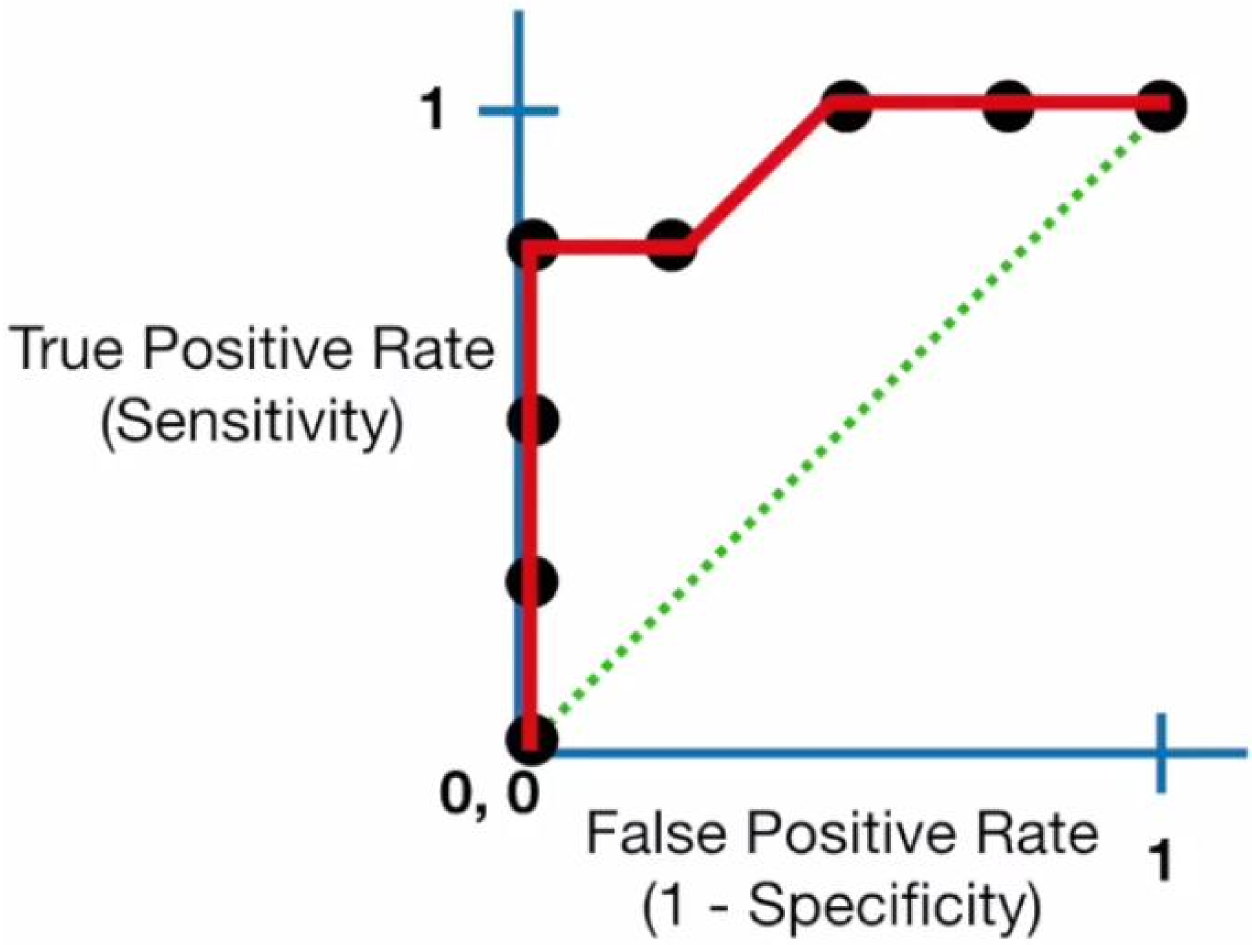


ROC Curves (cont.)

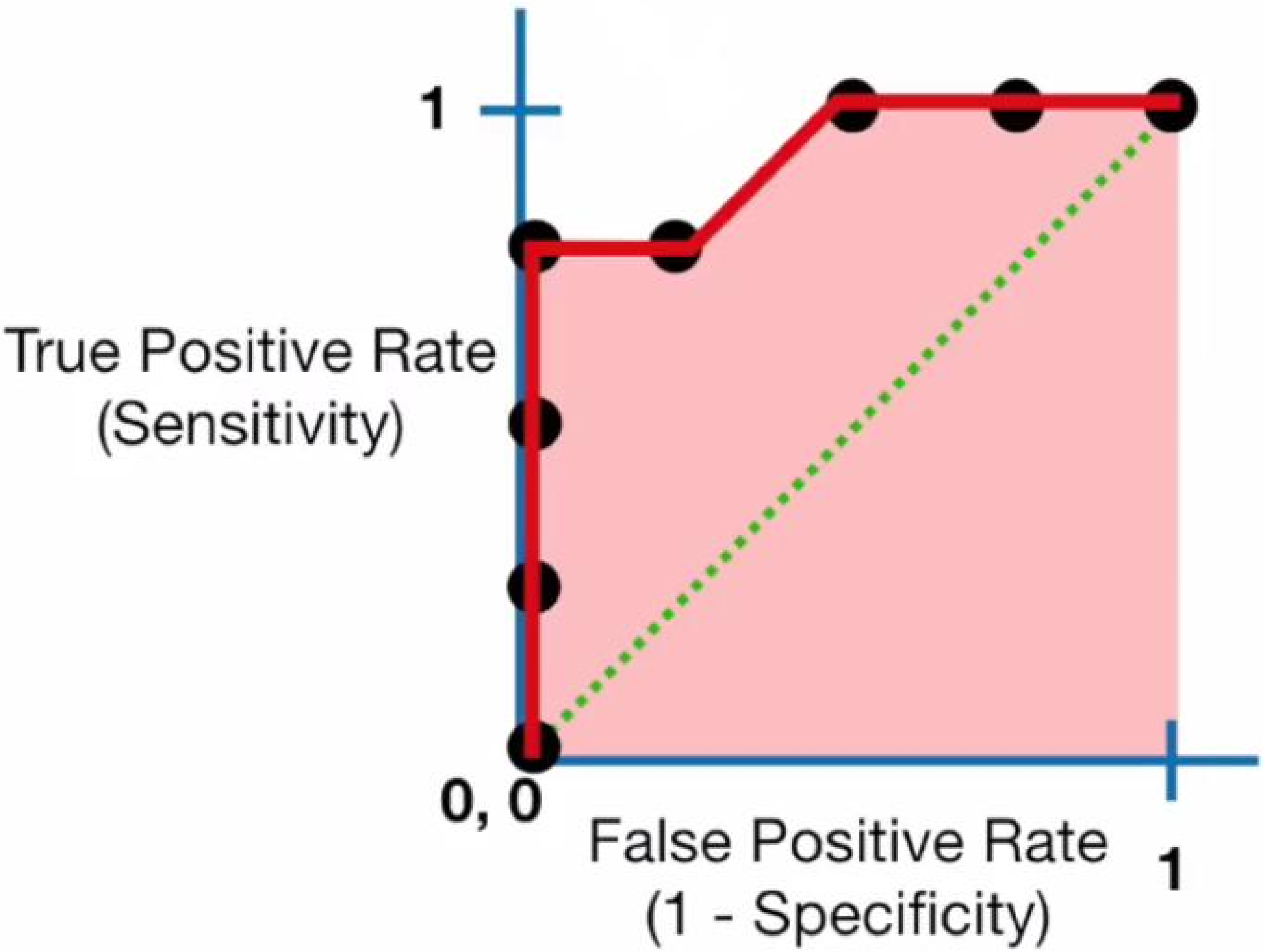


ROC Curves (cont.)

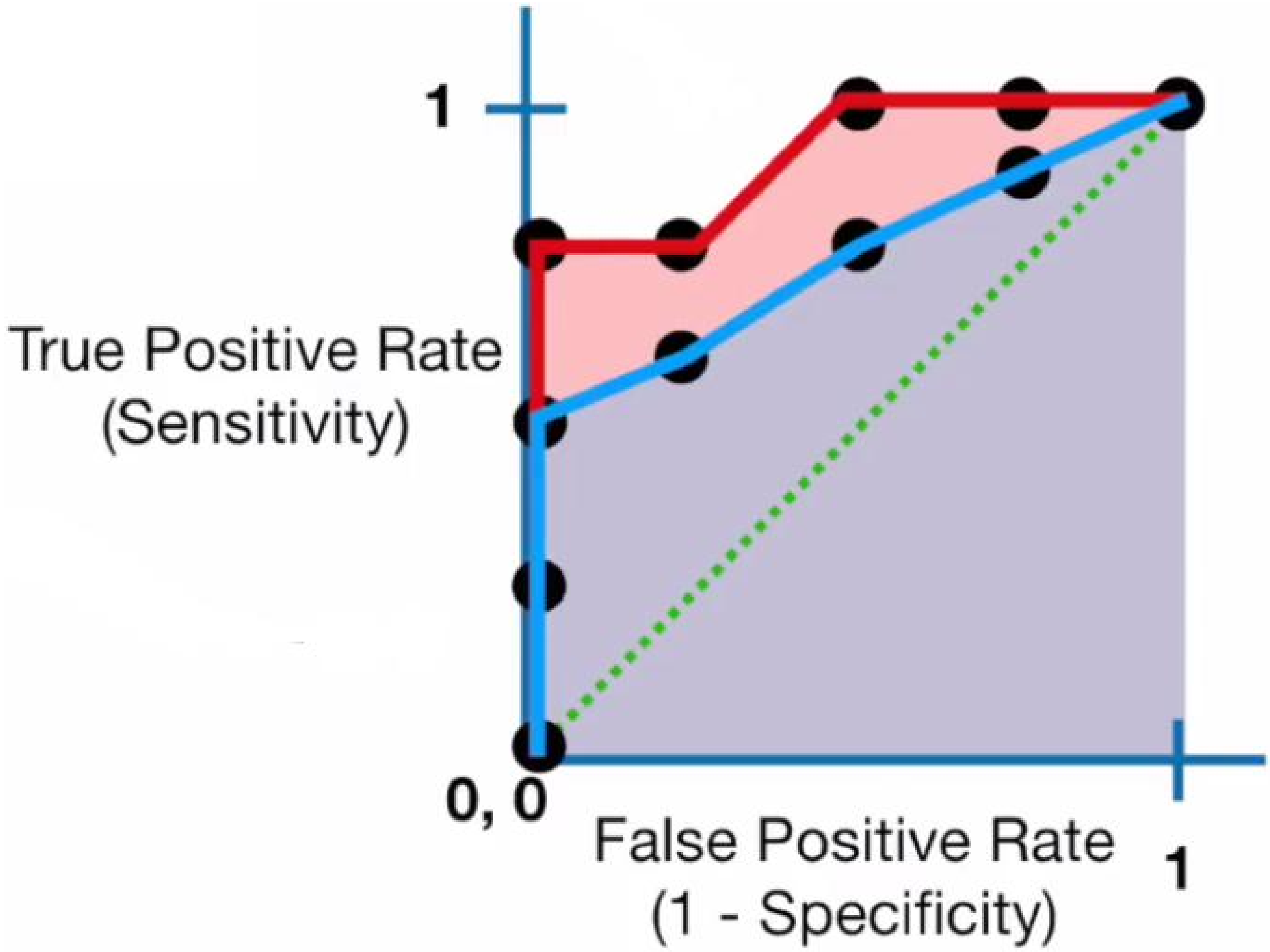
- How to determine which threshold is the best?



Area Under Curve (AUC)



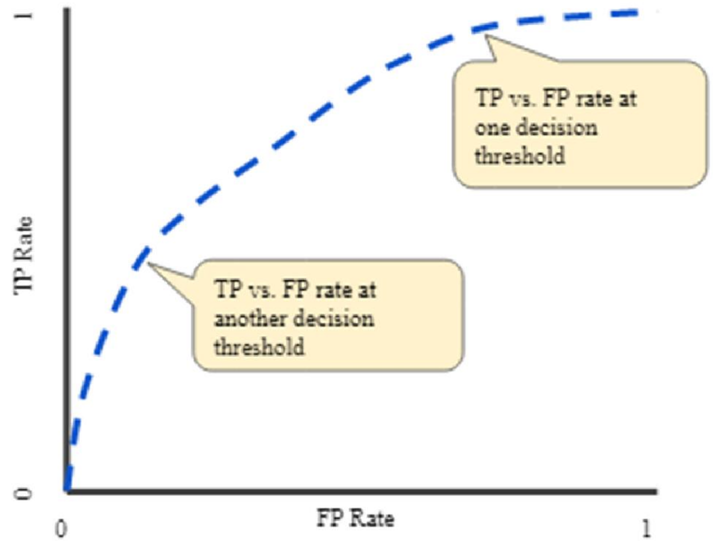
Comparing Different Classifiers



Summary

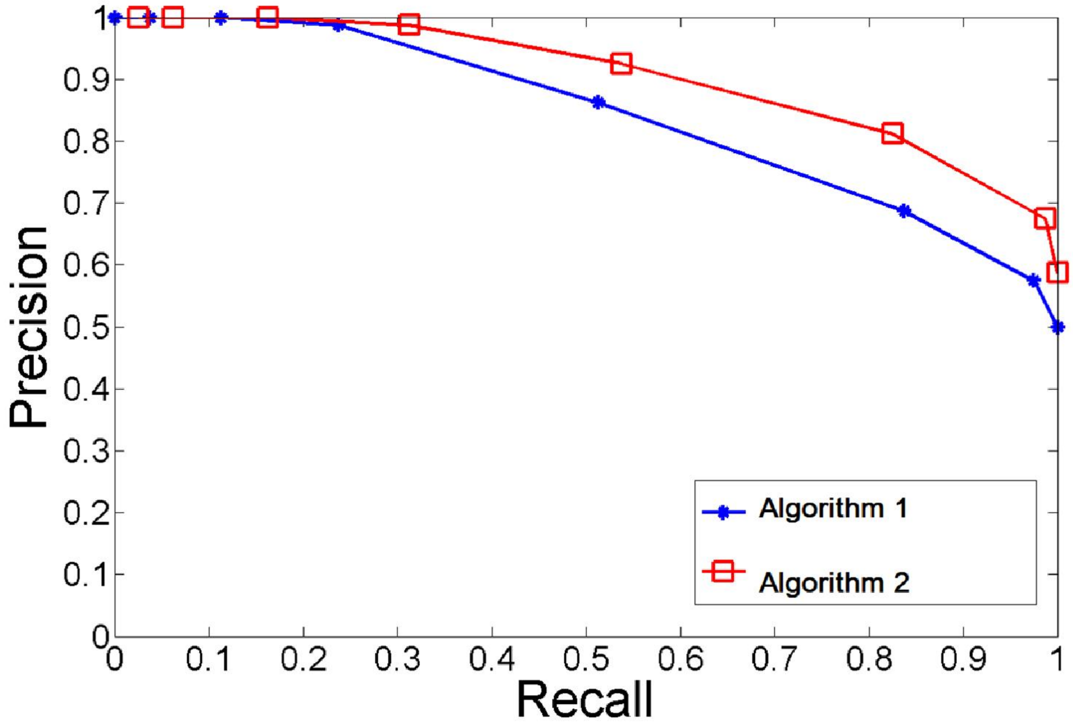
- An ROC curve (receiver operator characteristic curve) is a graph showing the performance of a classification model at all classification thresholds.

- It is worth noting that even though the example in this lecture was based on logistic regression (LR), ROC curves are not limited to LR and apply to all types of classifiers.



Further Reading

- The Precision-Recall Curve is more informative than ROC curve when dataset is imbalanced. Why?



References

- Logistic Regression and ROC lectures on StatQuest

- Tom Mitchell, Generative and Discriminative Classifiers: Naïve Bayes and Logistic Regression, Machine Learning (2nd ed.), Chapter 3, McGraw Hill, 2015 (last visit!!).