

ML EVALUATION

**In this session**

* ML train and evaluation circle
* How to read Histogram
* How to read Box Plot
* Adding Evaluate Model
* How to read ROC curve
* Area Under the Curve (AUC)
* How to read Evaluation metrics

ML evaluation circle

Data ingest

No

Need

improve?

Deploy

Yes

Evaluate

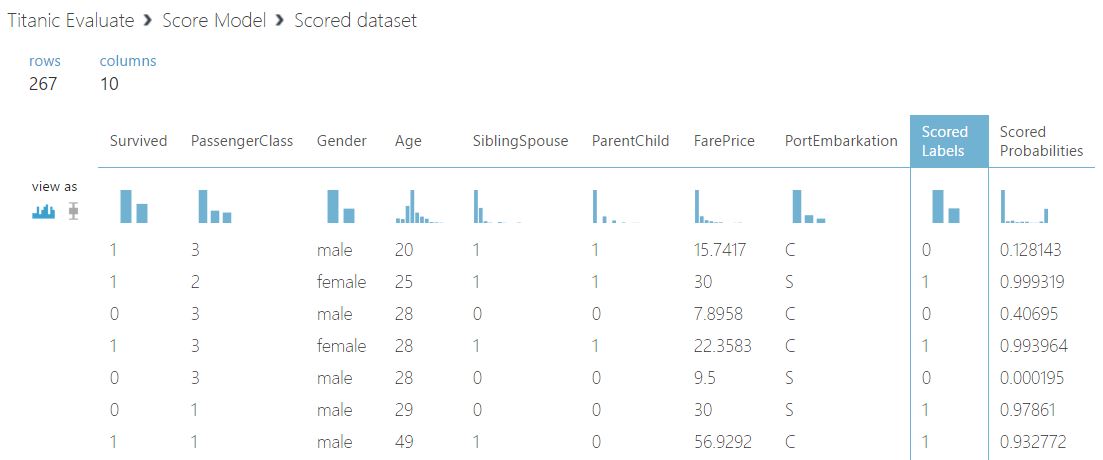
Data Cleansing

Feature Engineering

Train

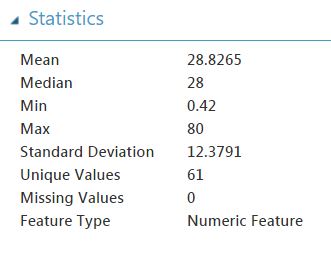
Score

How to read Scoring results



* This table = Scored dataset
* Row = 267 / Columns = 10
* Total column = 10 / Left 8 = features / Right 2 = prediction results
* Scored Label 0 = dead 1 = survived
* Scored Probabilities (SP) SP <=0.5 == dead / SP > 0.5 == survived

How to read Scoring Statistics

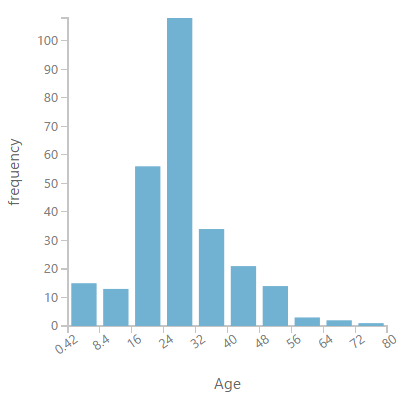


Show Statistics of the Scored dataset

* Mean = Sum of all the values divided by the number of values
* Median = The midpoint of the data after being ranked
* Standard Deviation = The square root of the variance
* Unique Values
* Missing Value

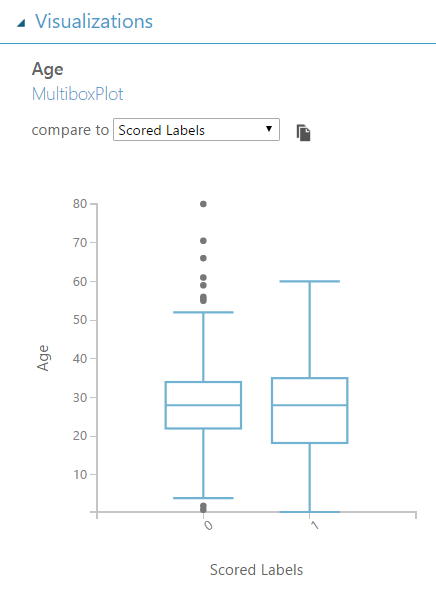
How to read Score Histogram

**Histogram**

* Representation: distribution of numerical data
* Bin: series of intervals (bin)
* Count: values fall into each interval

How to read Box Plot

**Box Plot**

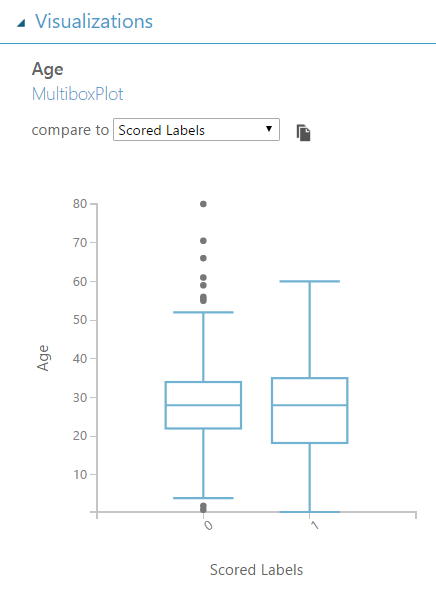


Box Plot (whisker) is a standardized way of displaying the distribution of data

* Median: marks the mid-point of the data
* Box: middle 50% of scores for the group.
* Upper quartile: 75% of the scores fall below the upper quartile.
* Lower quartile: 25% of scores fall below the lower quartile.
* Whiskers: scores outside the middle 50%

0 = dead

Box Plot Definitions



Extremes

Outliers

Upper whisker

25% Quartile group 4

Upper Quartile

Middle Quartile / median

25% Quartile group 3

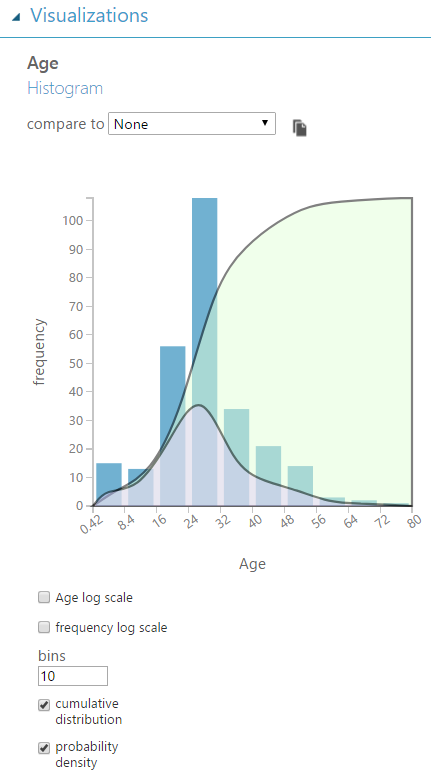
25% Quartile group 2

Lower Quartile

25% Quartile group 1

Lower whisker

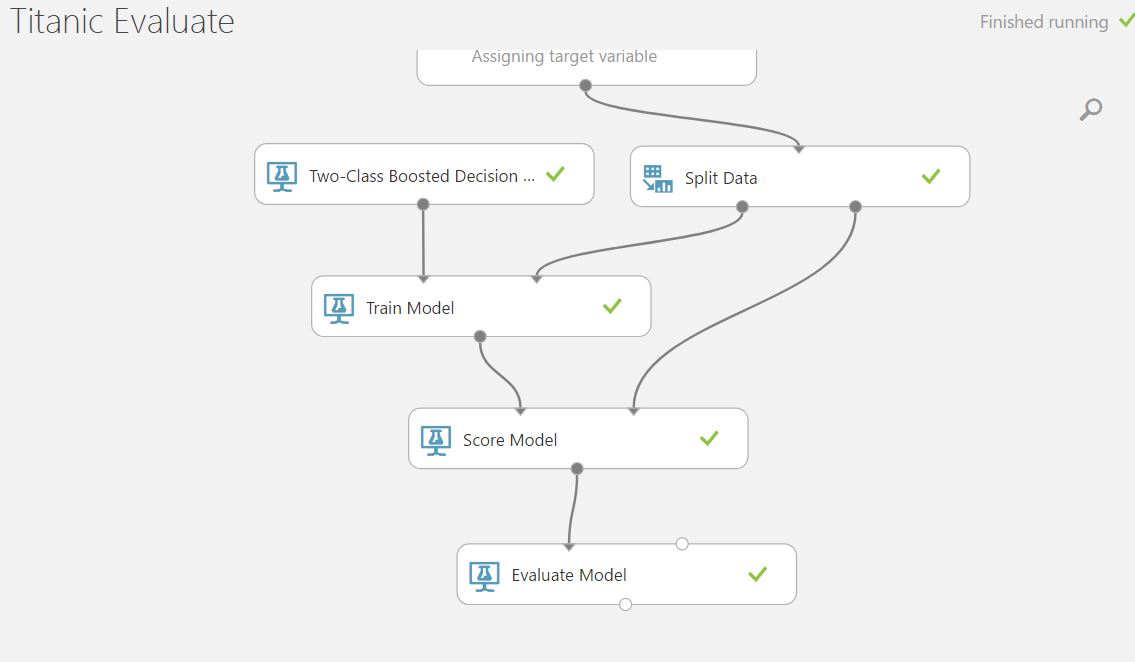
Histogram option

**Histogram options**

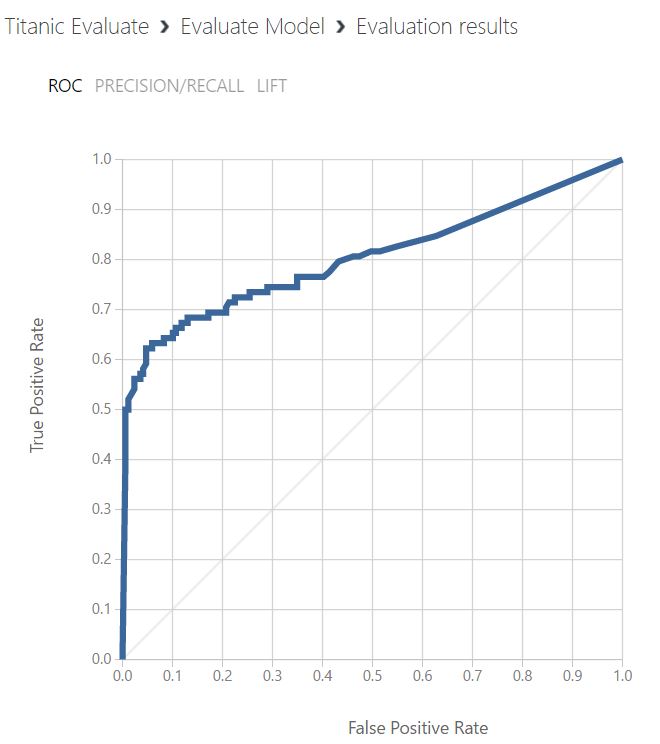
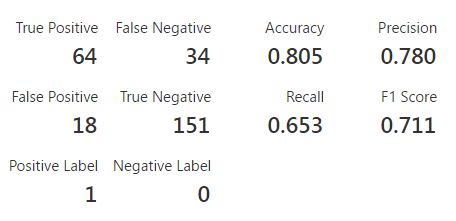
* Cumulative distribution function (cdf): shows "How common are samples that are *less than or equal* to this value?"
* Probability density function (pdf): shows "How common are samples at exactly this value?"
* Scale: scaling the distribution
* bins: number of bin

Adding Evaluate Model

Adding Evaluate Model

1. Open Titanic 1 Experiment
2. Save as Titanic Evaluate
3. Add Evaluate Model
4. Run the Experiment

Receiver Operating Characteristic (ROC) Curve



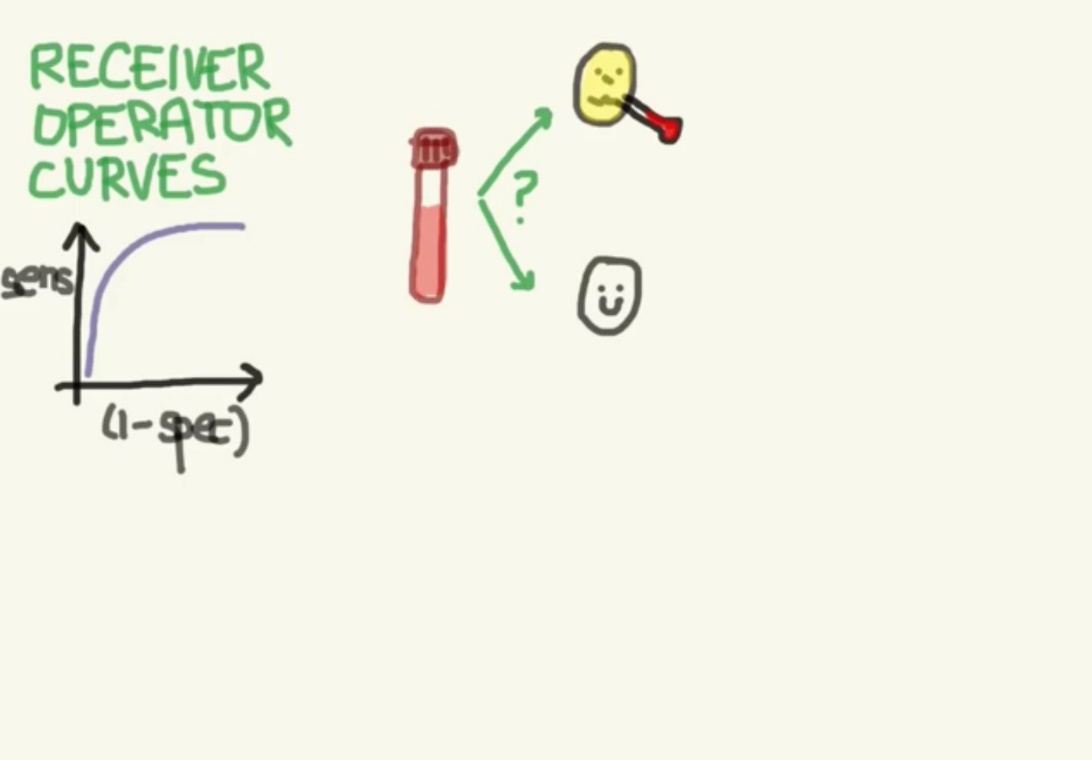


True Positive Rate (TPR)

False Positive Rate (FPR)

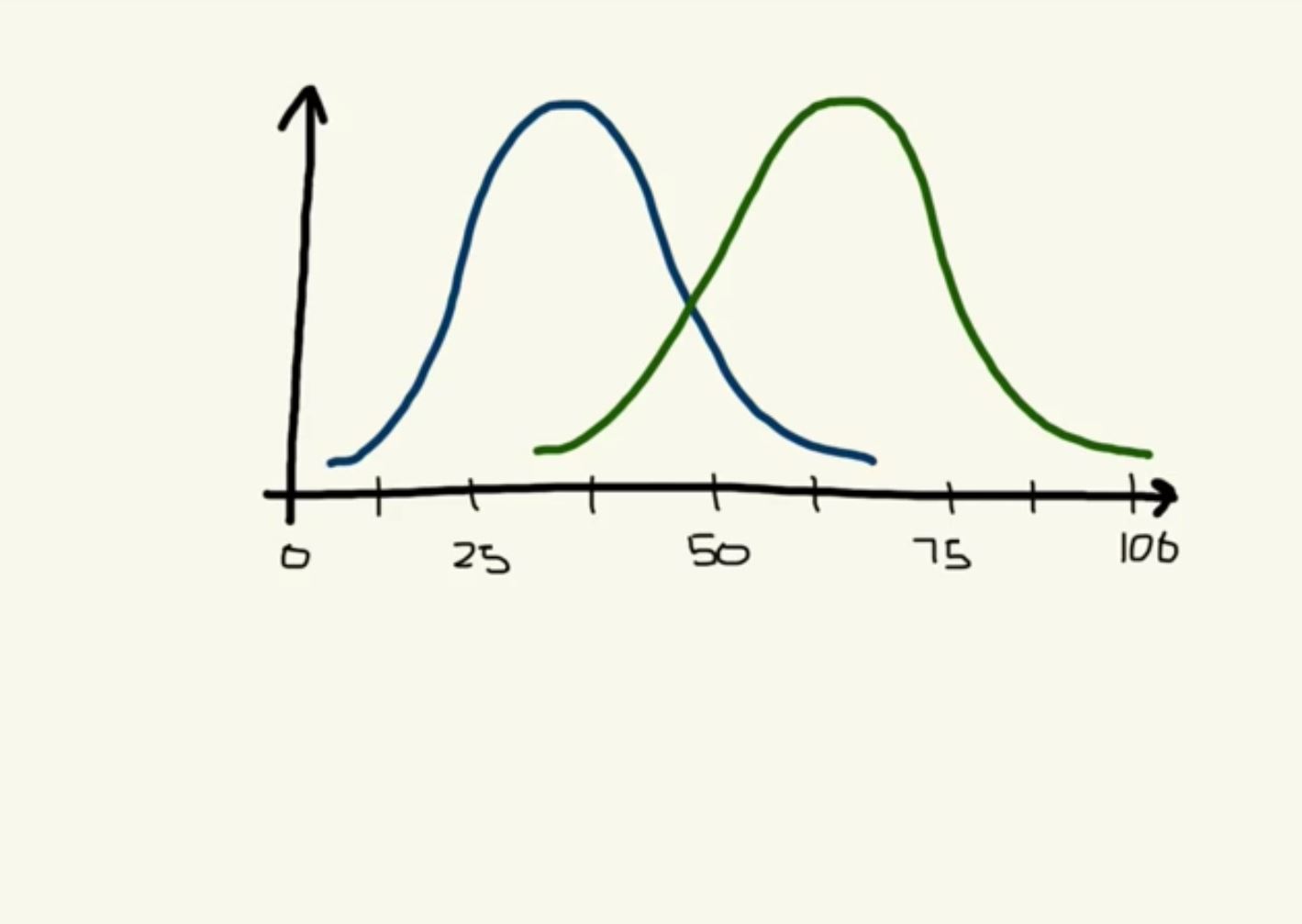
How to read ROC curve

ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.



ROC curve prediction result who have disease who don’t

Distribution score



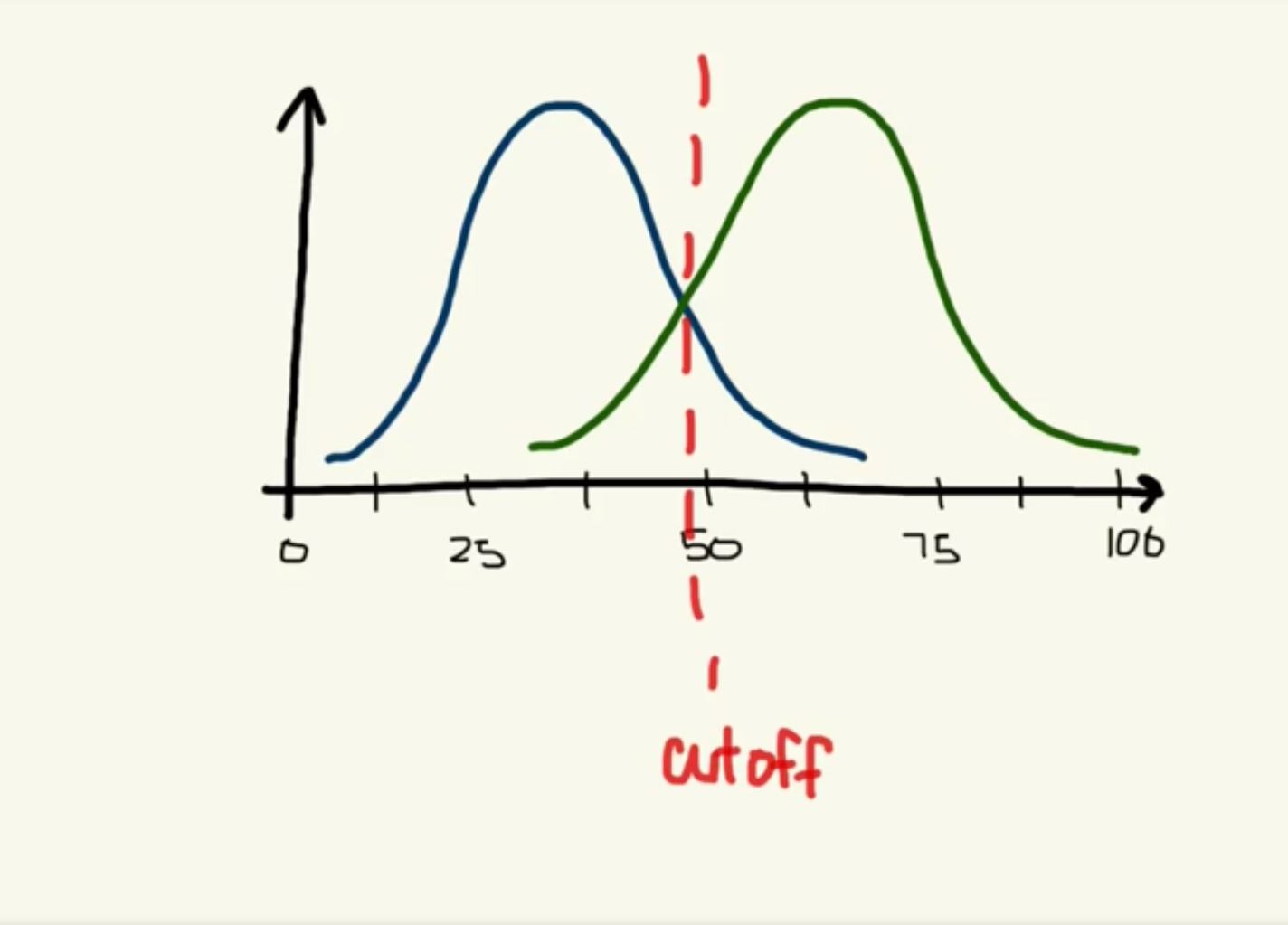
Y

X

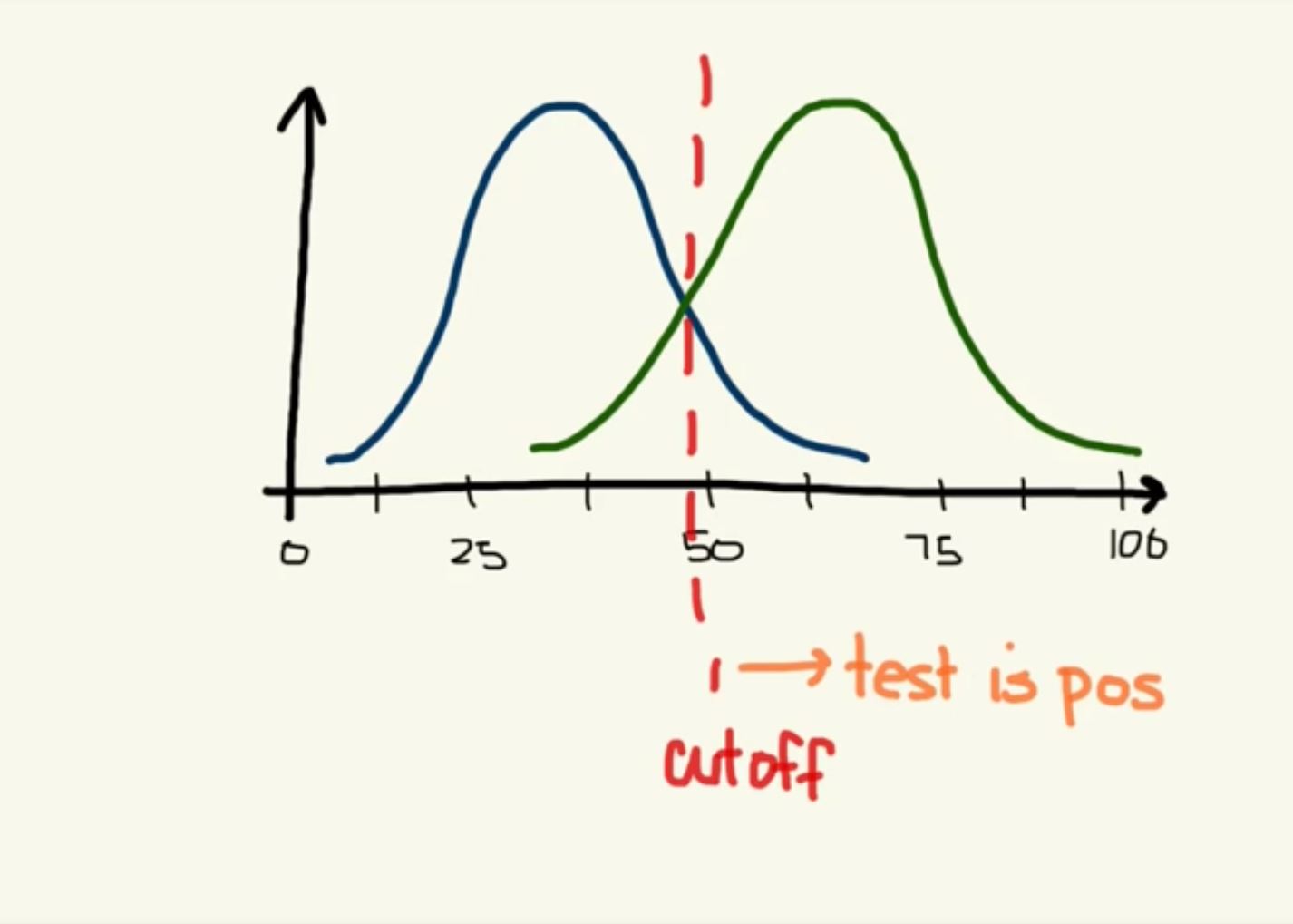
Left distribution = patient who do NOT have disease (survived) / Right = have disease (dead)

x axis = score / y axis = number of patient

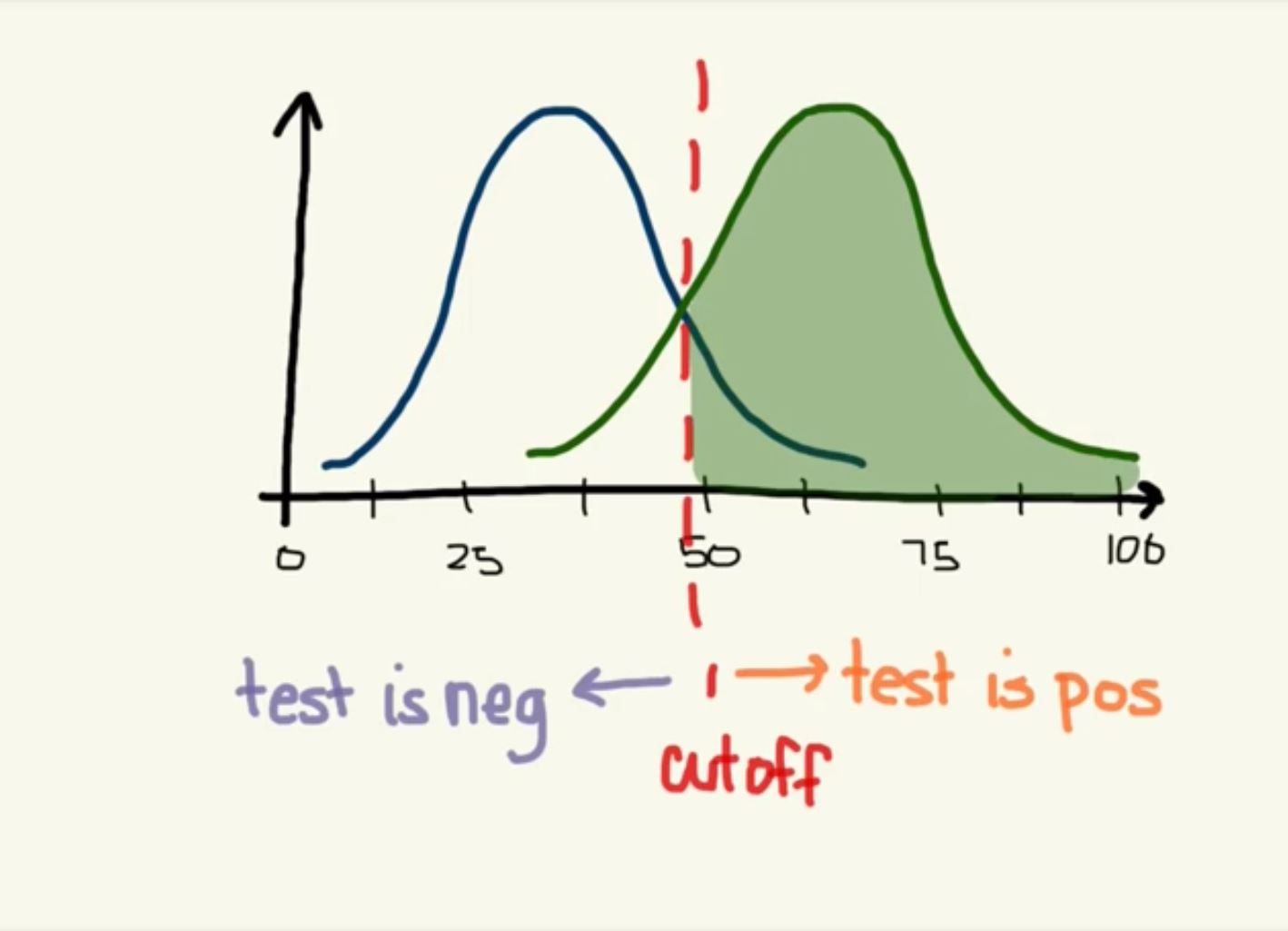
Cutoff line



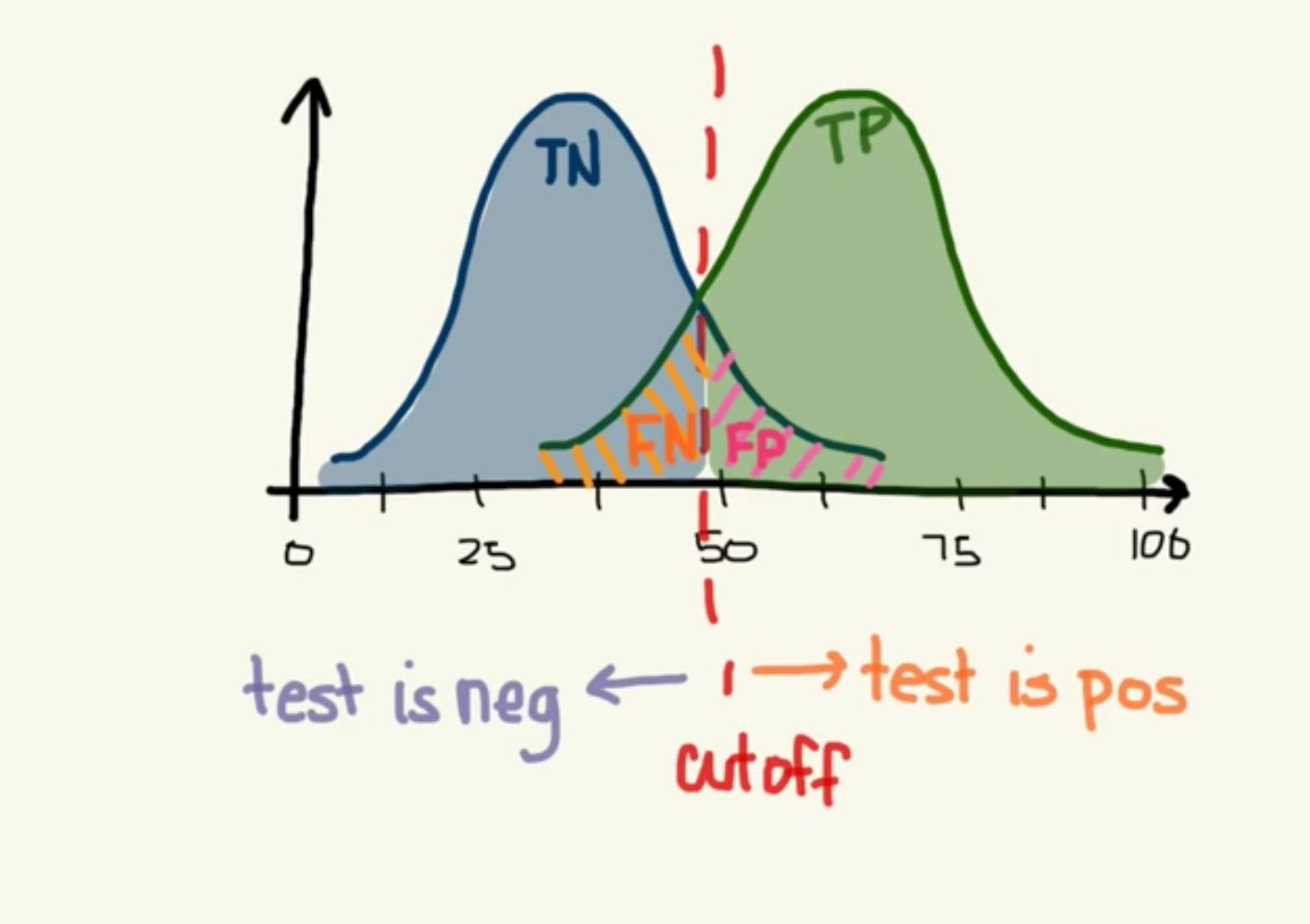
Area where the test is positive



Area where the test is negative



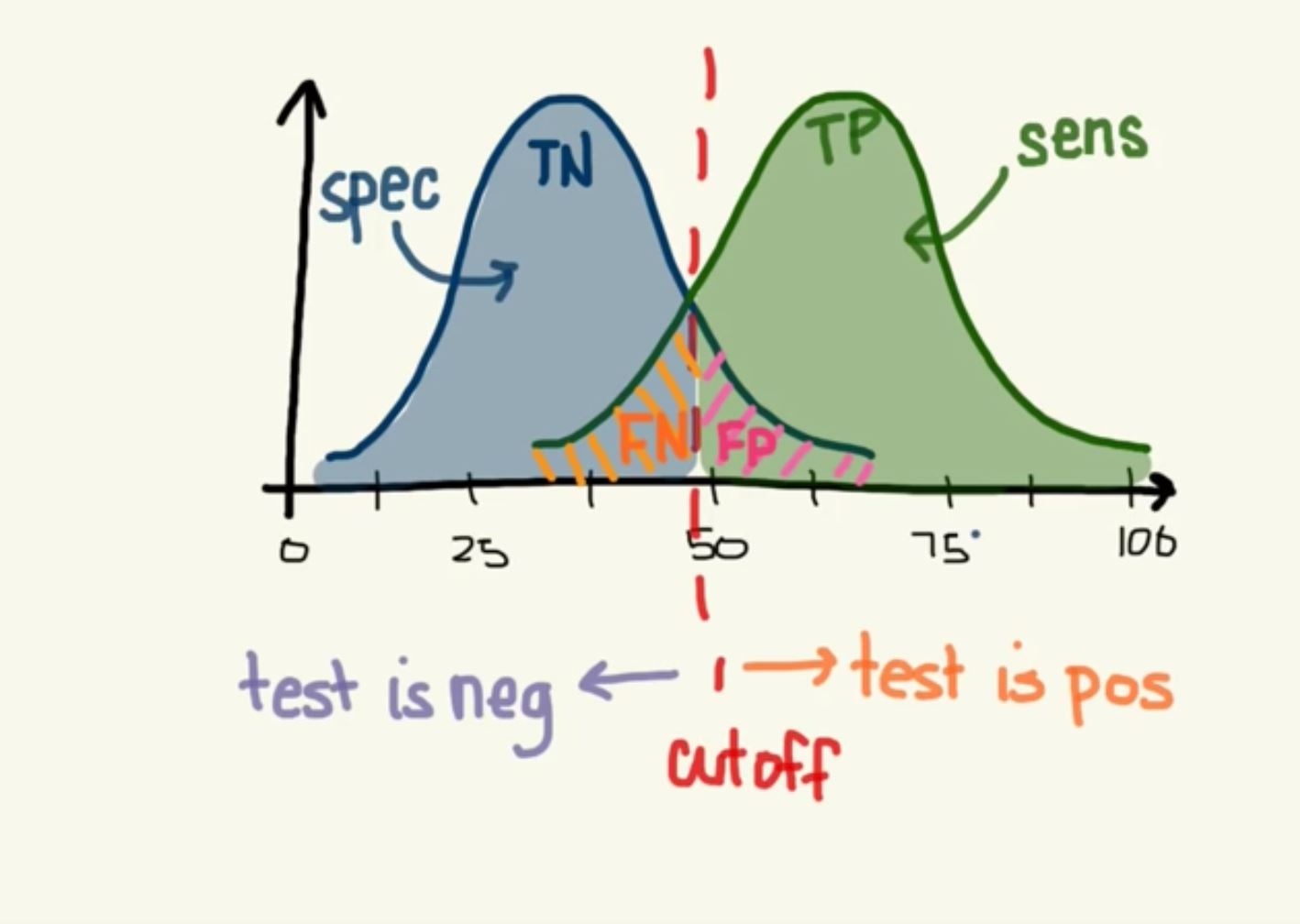
True Negative (TN), False Negative (FN) / True Positive (TP), False Positive (FP)



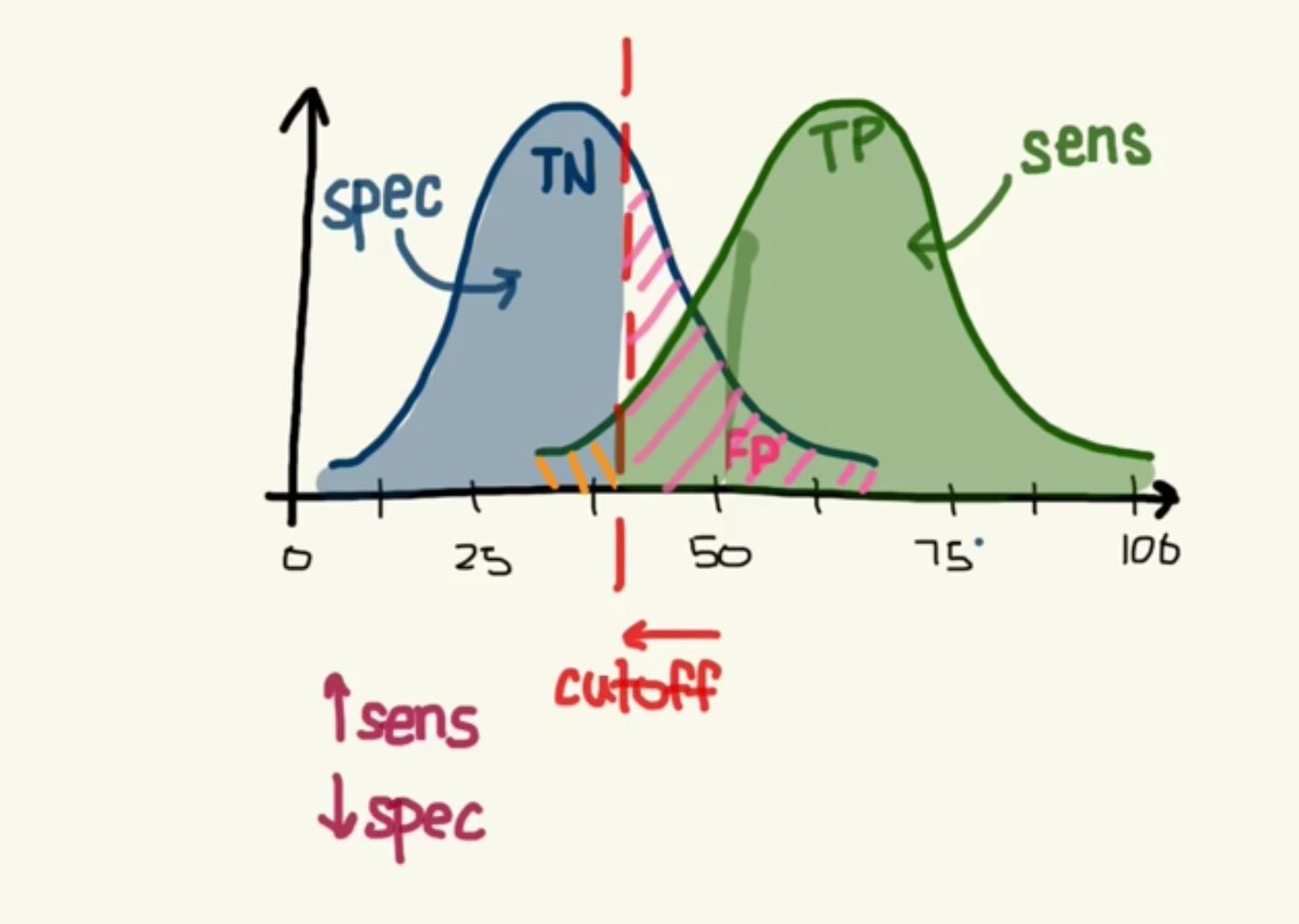
ROC Specificity / Sensitivity

Specificity = True Negative Rate

Sensitivity (Recall) = True Positive Rate



Move cutoff to the left Sens++ / Spec--



Move cutoff to right Sens-- / Spec++

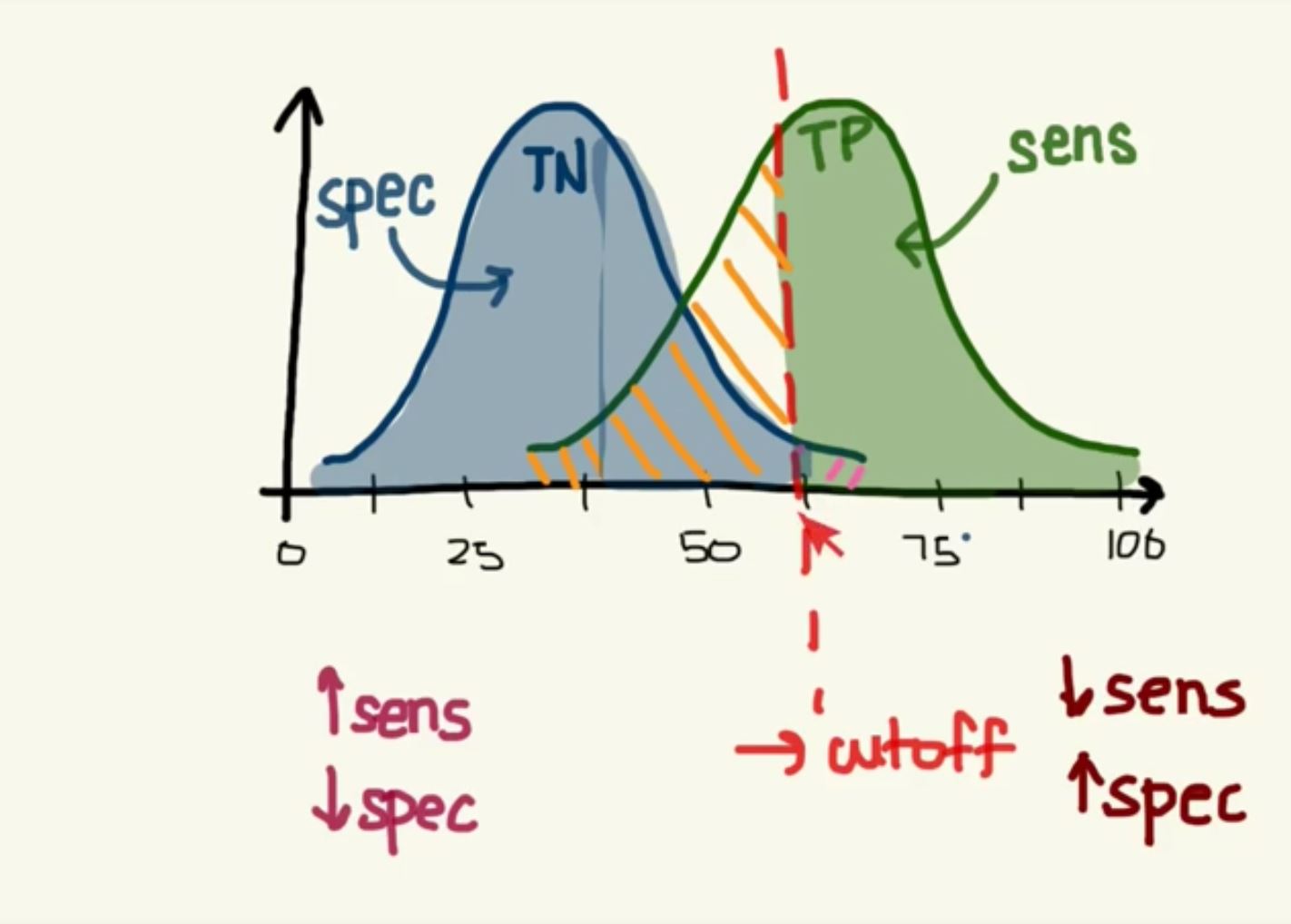
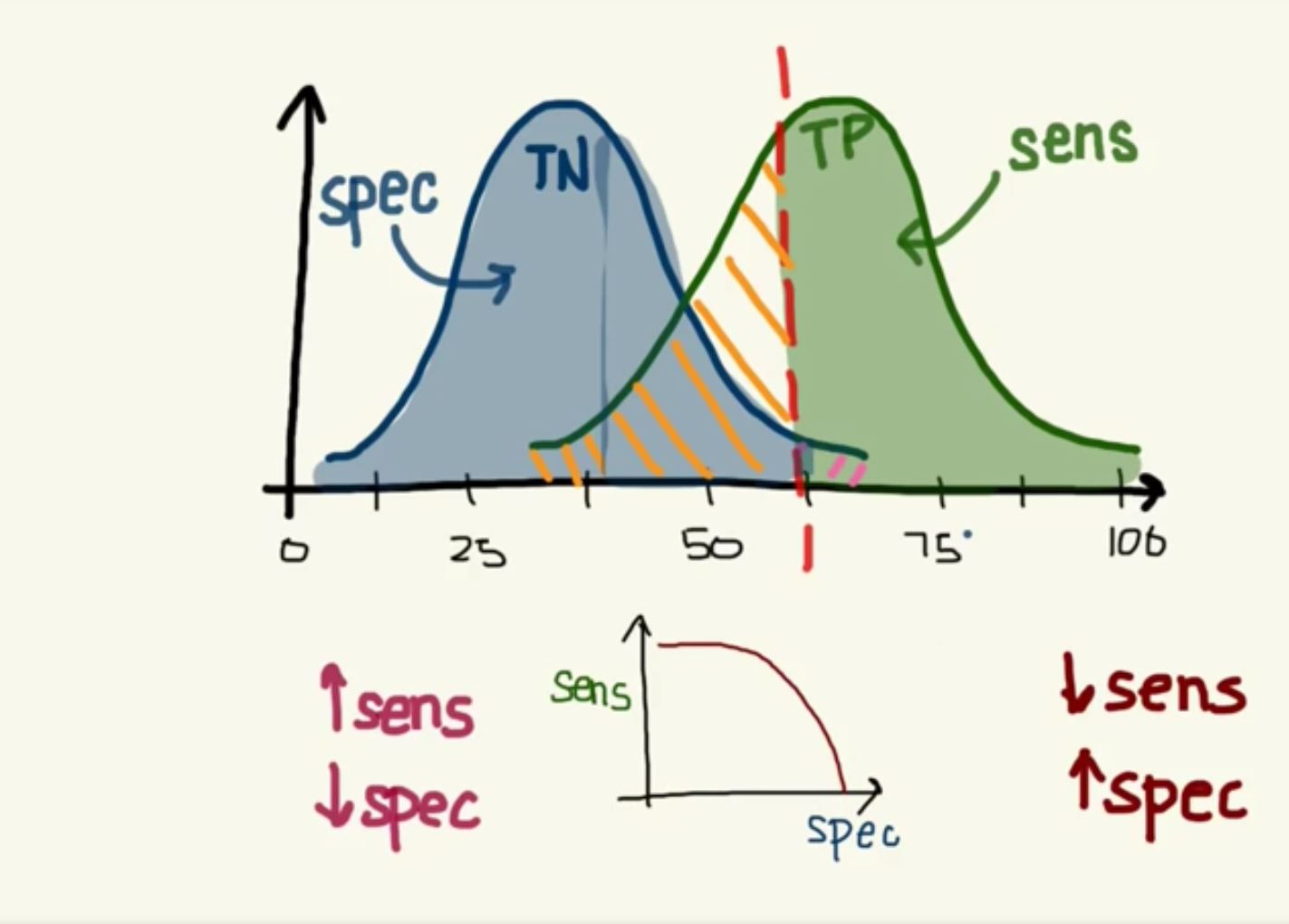
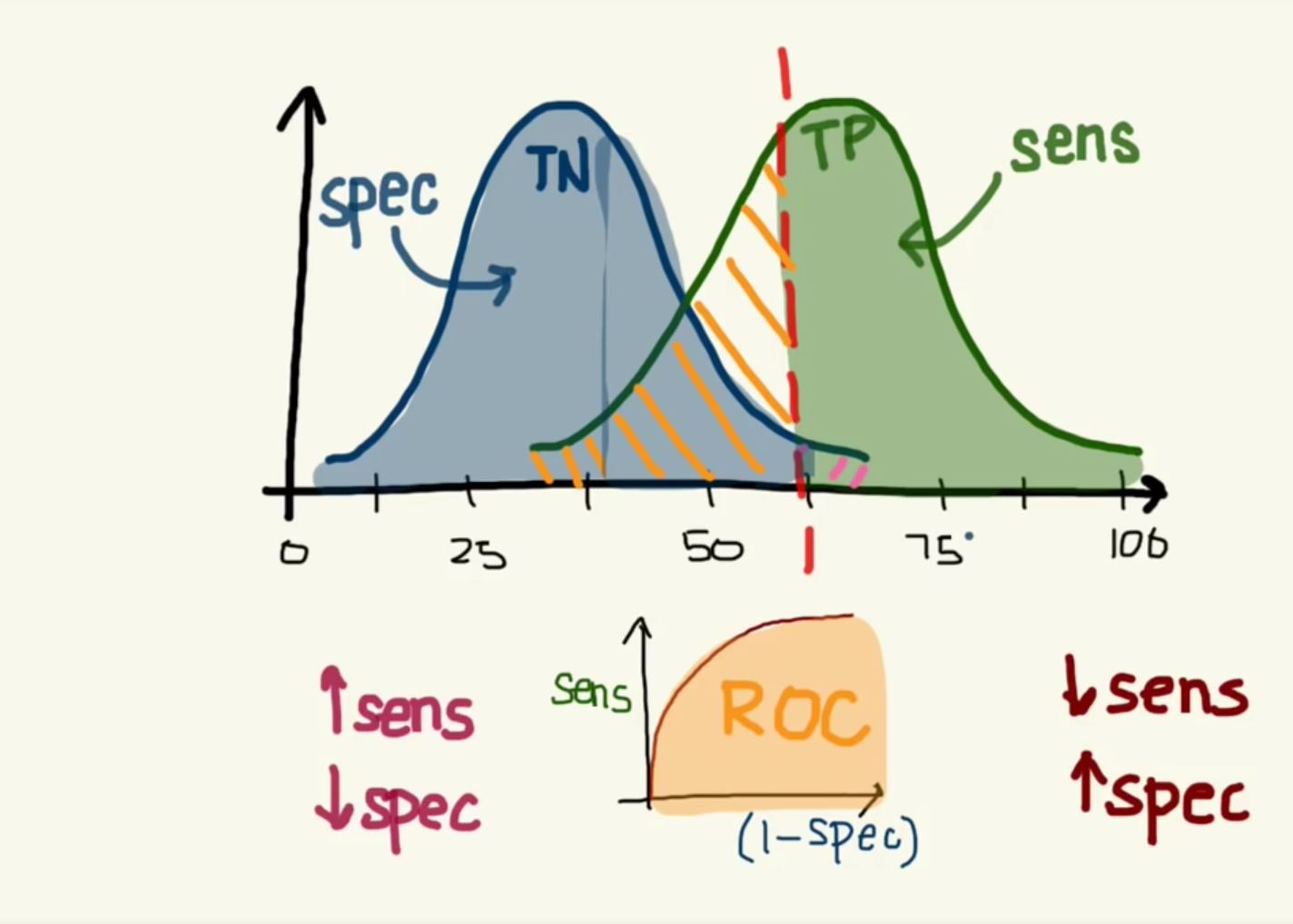


Chart proportion of Sens / Spec

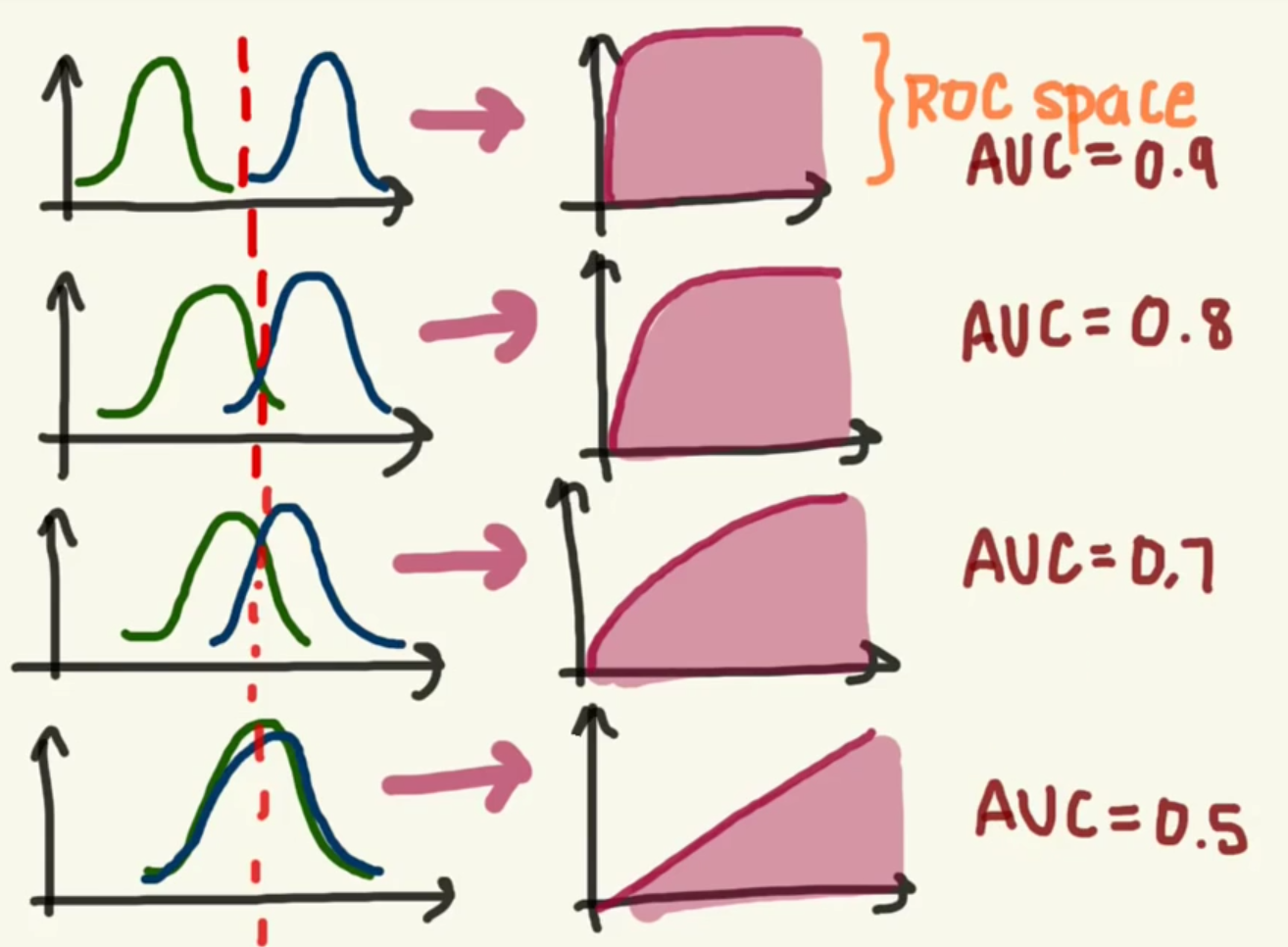


ROC curve = proportion of Sens / (1 – Spec)

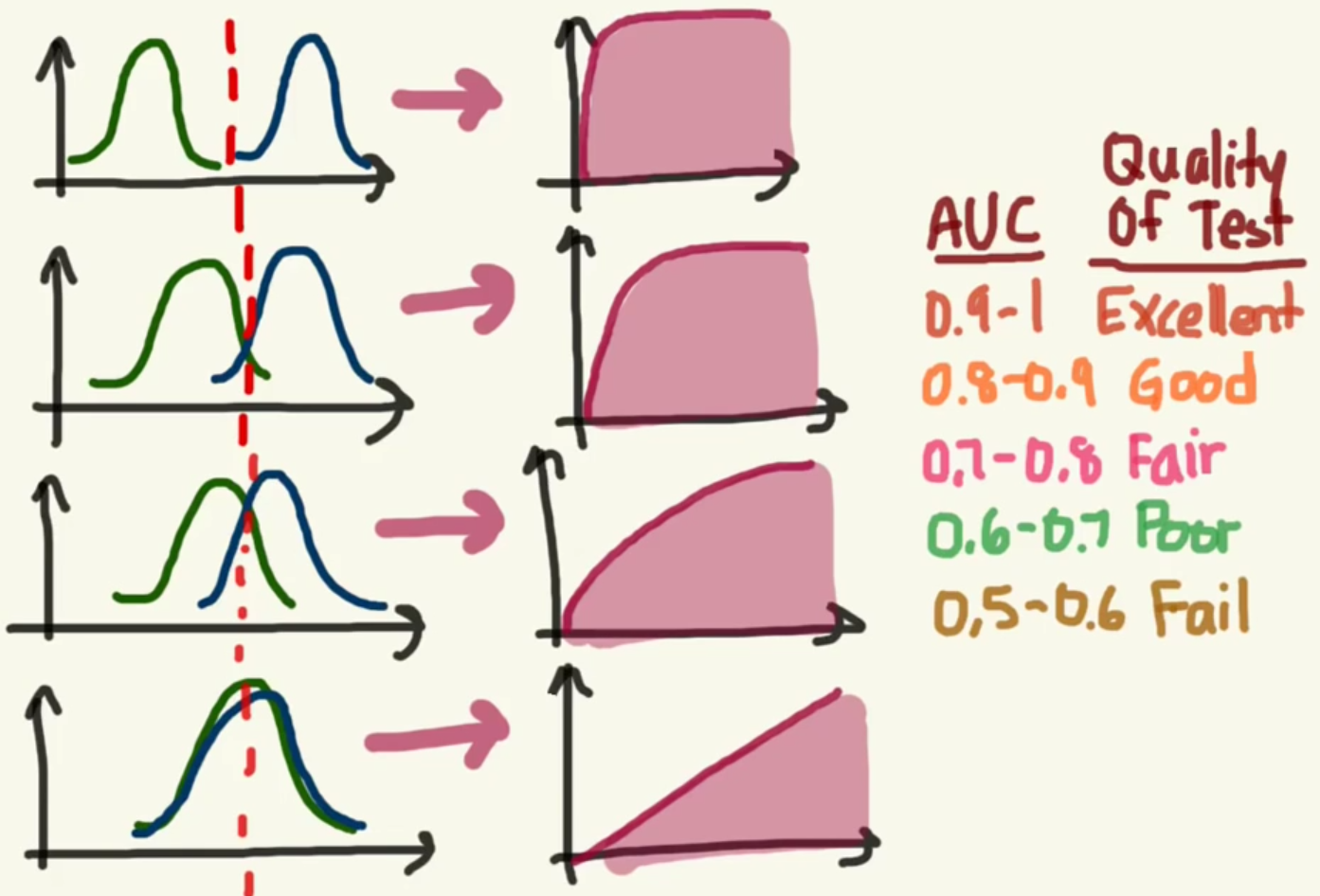


Area Under the Curve (AUC)

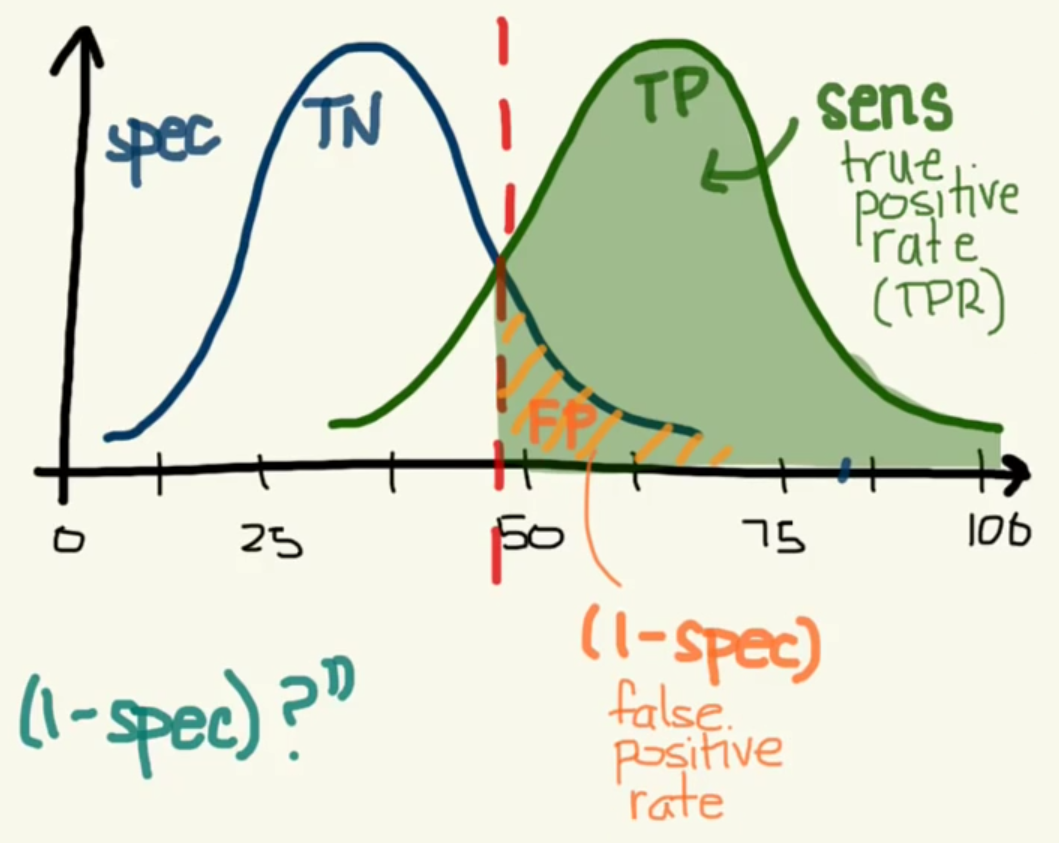
AUC is used to determine which of the used models predicts the classes best.



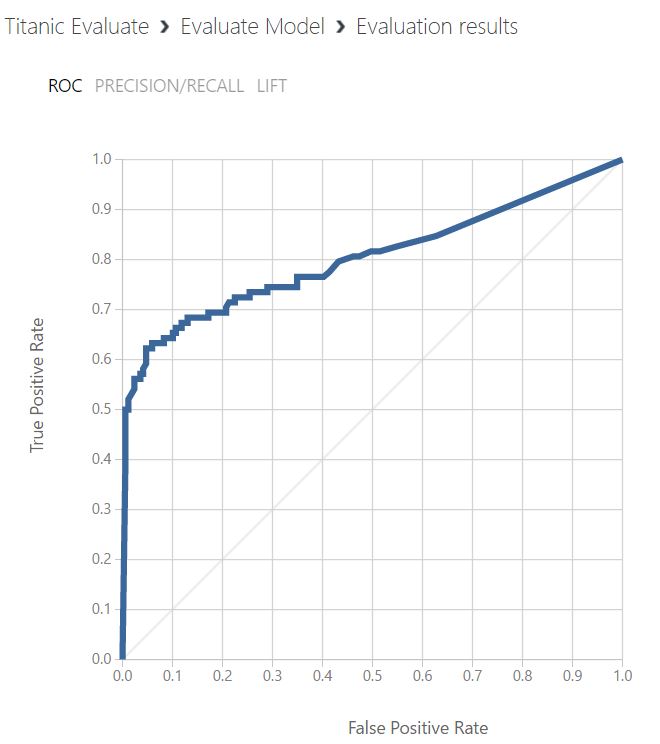
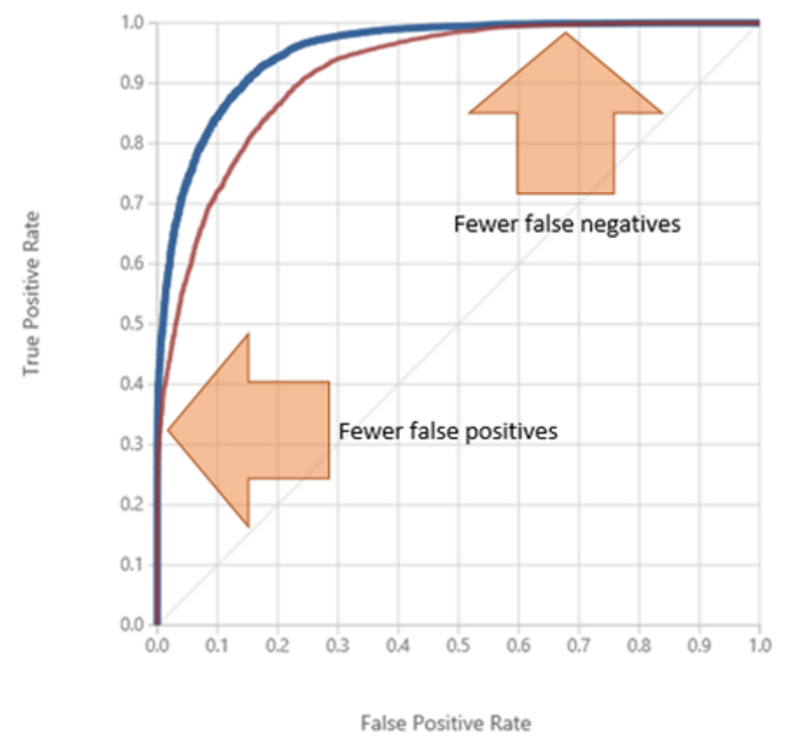
AUC score



What is (1-spec)

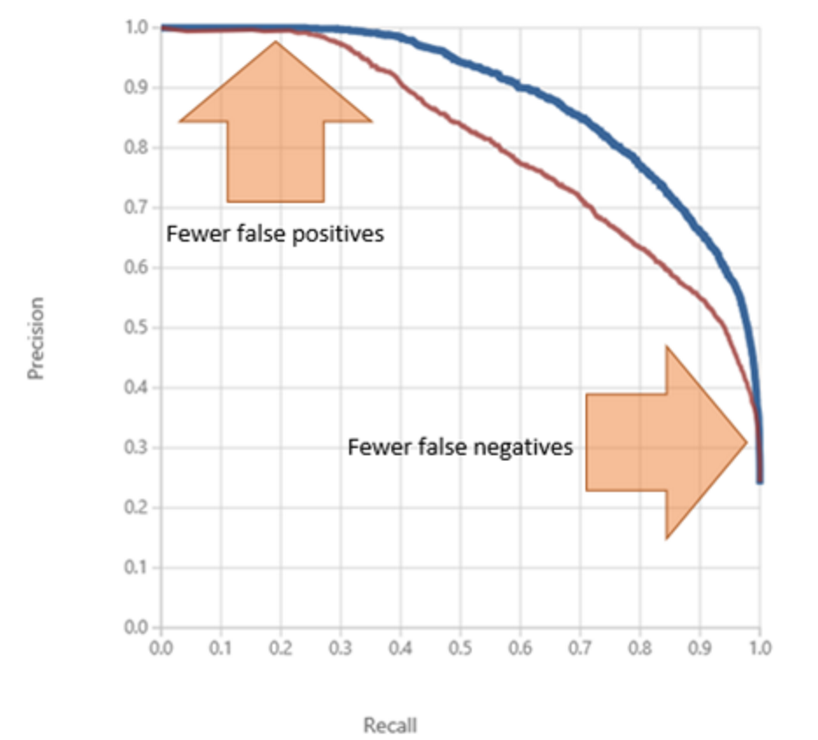
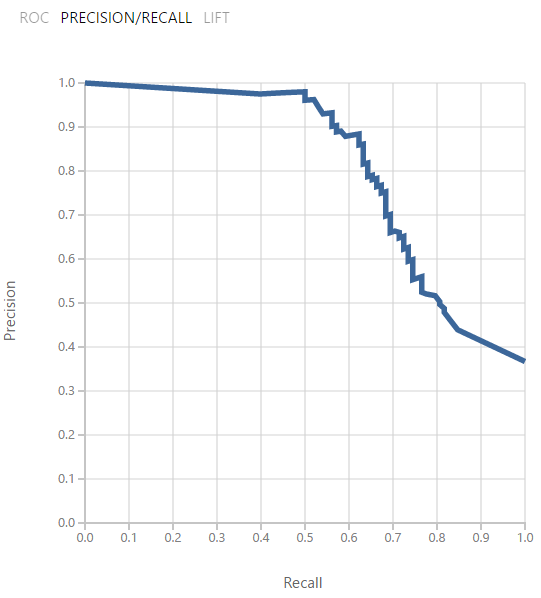


ROC Curve



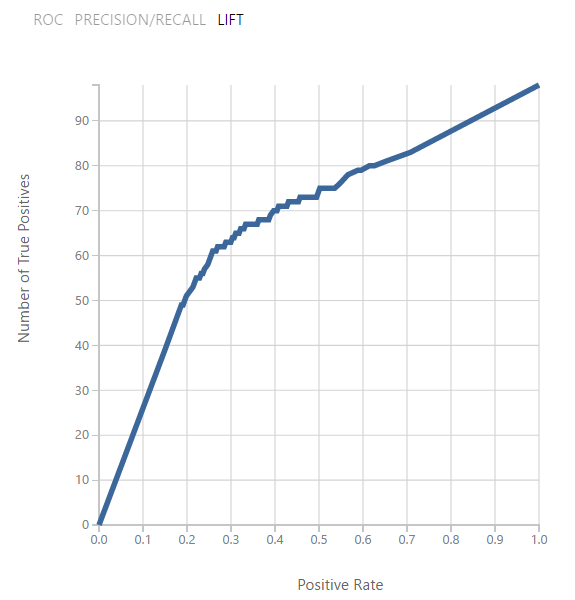
ROC curve displays the fraction of true positives out of the total actual positives. The higher and further to the left, the more accurate the model is. As you do experiments you want to see the curve move higher and to the left.

PRECISION/RECALL



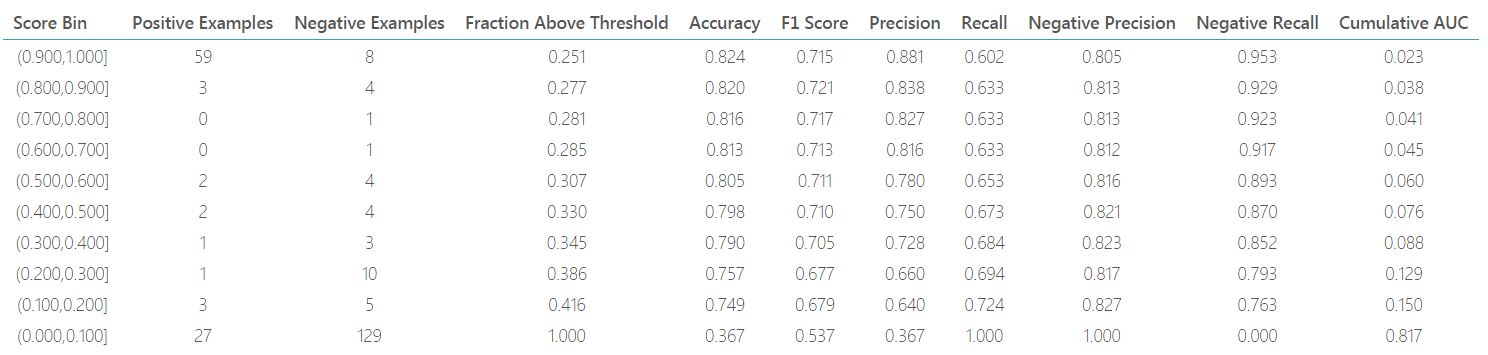
Precision represents the fraction of retrieved instances that are relevant, whereas recall represents the fraction of relevant instances that are retrieved. The “sweet spot” for the ideal model is in the upper right corner

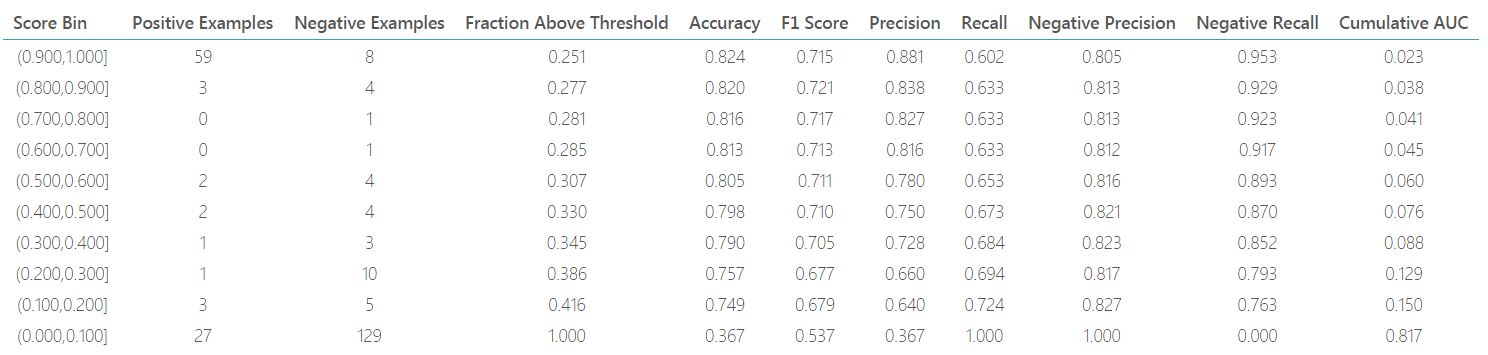
LIFT curve



Lift curve is a variation on the ROC curve. It measures the fraction of true positives, in relation to the target response probability.

Reading Evaluation metrics

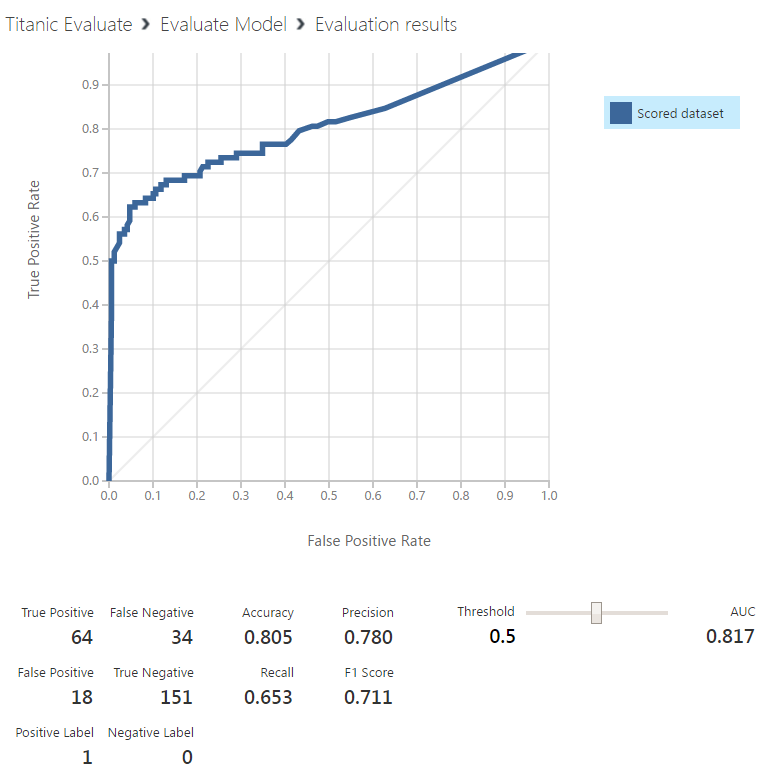




Evaluation metrics variable

* True Positive (TP): Correctly identified e.g. Sick people correctly diagnosed as sick
* False Positive (FP): Incorrectly identified e.g. healthy people incorrectly identified as sick
* True Negative (TN): Correctly rejected e.g. healthy people correctly identified as healthy
* False Negative (FN): Incorrectly rejected e.g. Sick people incorrectly identified as healthy
* Accuracy : The proportion of the total number of predictions that is correct. (TP + TN) / (TP + TN + FP + FN)
* Precision: is the proportion of positive cases that were correctly identified. TP / (TP + FP)
* Recall: Sensitivity or Recall is the proportion of actual positive cases which are correctly identified. TP / (TP + FN)
* F1 Score: is the harmonic mean of precision and Recall. 2TP / (2TP + FP + FN)
* Threshold: Threshold is the value above which it belongs to first class and all other values to the second class. E.g. if the threshold is 0.5 then any patient scored more than or equal to 0.5 is identified as sick else healthy.

Titanic evaluation results



* Positive Label: 1 = survived
* Negative Label: 0 = dead
* True Positive: correctly predict survived
* True Negative: correctly predict dead
* False Positive: incorrectly predict survived
* False Negative: incorrectly predict dead

More information

How to evaluate model performance in Azure Machine Learning

https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-evaluate-model-performance

This experiment ML model

Adding Evaluation model

<https://gallery.cortanaintelligence.com/Experiment/Titanic-1-2>

Adding "Two-Class Decision Forest"

https://gallery.cortanaintelligence.com/Experiment/Titanic-compare-two-alogrithm