وزارة الاتصـــــــالات
وتكنولوجيا المعلومات

رواد مصر الرقمية

Logistic Regression with Scikit-Learn

Part One: Exploratory Data Analysis

رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

00-Logistic-Regression.ipynb

jupyter

Logistic Regression with Scikit-Learn

Part Two: Creating and Training a Model

00-Logistic-Regression.ipynb

Classification Performance Metrics

Part One: Confusion Matrix Basics

Classification Metrics

- You’ve probably heard of terms such as “false positive” or “false negative”. As well as metrics like “accuracy”.
- But what do these terms actually mean mathematically?

Classification Metrics

- Imagine we've developed a test or model to detect presence of a virus infection in a person based on some biological feature.
- We could treat this as a Logistic Regression, predicting:
    - 0 - Not Infected (Tests Negative)
    - 1 - Infected (Tests Positive)

Classification Metrics

- It is unlikely our model will perform perfectly. This means there 4 possible outcomes:
  - Infected person tests positive.
  - Healthy person tests negative.

Classification Metrics

- It is unlikely our model will perform perfectly. This means there 4 possible outcomes:
    - Infected person tests positive.
    - Healthy person tests negative.
        - *Note, these are the outcomes we want! But it is unlikely our test is perfect...*

Classification Metrics

- It is unlikely our model will perform perfectly. This means there 4 possible outcomes:
  - Infected person tests positive.
  - Healthy person tests negative.
  - Infected person tests negative.
  - Healthy person tests positive.

Classification Metrics

- Based off these 4 possibilities, there are many error metrics we can calculate.
- First, let's start by visualizing these four possibilities as a matrix.

Classification Metrics

- Confusion Matrix

	ACTUAL	
	INFECTED	HEALTHY

- ## Confusion Matrix

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED		
	HEALTHY		

- Confusion Matrix

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	TRUE POSITIVE	
	HEALTHY		

Classification Metrics

- Confusion Matrix

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	TRUE POSITIVE	
	HEALTHY		TRUE NEGATIVE

Classification Metrics

- Confusion Matrix

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	TRUE POSITIVE	FALSE POSITIVE
	HEALTHY		TRUE NEGATIVE

Classification Metrics

- Confusion Matrix

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	TRUE POSITIVE	FALSE POSITIVE
	HEALTHY	FALSE NEGATIVE	TRUE NEGATIVE

Classification Metrics

- Imagine a test group of 100 people:

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED		
	HEALTHY		

Classification Metrics

- 5 are infected. 95 are healthy.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED		
	HEALTHY		

Classification Metrics

- We tested all of them with these results:

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Classification Metrics

- What is accuracy?

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Classification Metrics

- What is accuracy?

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

- Accuracy:
  - How often is the model correct?

Acc = (TP+TN)/Total

Classification Metrics

- Calculating accuracy:

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

(4+93)/100 = 97% Accuracy

- Accuracy:
  - How often is the model correct?

Acc = (TP+TN)/Total

Classification Metrics

- Is this a good value for accuracy?

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

(4+93)/100 = 97% Accuracy

- Accuracy:
  - How often is the model correct?

Acc = (TP+TN)/Total

Classification Metrics

- The accuracy paradox...

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

(4+93)/100 = 97% Accuracy

- Accuracy:
  - How often is the model correct?

Acc = (TP+TN)/Total

Classification Metrics

- Imagine we **always** report back "healthy"

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Classification Metrics

- Imagine we **always** report back "healthy"

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	0	0
	HEALTHY	5	95

(0+95)/100 = 95% Accuracy

- Accuracy:
  - How often is the model correct?

95% accuracy for a model that always returns "healthy"!

Classification Metrics

- **You may be thinking,** *"The numbers here are arbitrary, we just happen to get good accuracy in this made up case. Real world data would reflect poor accuracy if a model always returned the same result".*

Classification Metrics

- **You may be thinking,** *“The numbers here are arbitrary, we just happen to get good accuracy in this made up case. Real world data would reflect poor accuracy if a model always returned the same result”*.

Classification Metrics

- This is the accuracy paradox!
    - Any classifier dealing with **imbalanced** classes has to confront the issue of the accuracy paradox.
    - **Imbalanced** classes will always result in a distorted accuracy reflecting better performance than what is truly warranted.

Classification Metrics

- **Imbalanced** classes are often found in real world data sets.
  - Medical conditions can affect small portions of the population.
  - Fraud is not common (e.g. Real vs. Fraud credit card usage).

Classification Metrics

- If a class is only a small percentage (n%), then a classifier that always predicts the majority class will always have an accuracy of (1-n).
- In our previous example we saw infected were only 5% of the data.
- Allowing the accuracy to be 95%.

Classification Metrics

- This means we shouldn’t solely rely on accuracy as a metric!
- This is where precision,recall, and f1-score will come in.
- Let’s explore these other metrics in the next lecture.

رواد مصر الرقمية

وزارة الاتصالات
وتكنولوجيا المعلومات

00-Logistic-Regression.ipynb

jupyter

Classification Performance Metrics

Part Two: Precision and Recall

Classification Metrics

- We already know how to calculate accuracy and its associated paradox.
- Let's explore three more metrics that can help give a clearer picture of performance:
  - Recall (a.k.a. sensitivity)
  - Precision
  - F1-Score

Classification Metrics

- Let's begin with recall.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

- Recall:
  - When it actually is a positive case, how often is it correct?

(TP)/Total Actual Positives

Classification Metrics

- Let’s begin with recall.

ACTUAL

PREDICTED		INFECTED	HEALTHY
	INFECTED	4	2
	HEALTHY	1	93

Recall =
(TP)/Total Actual Positives

- Recall:
  - When it actually is a positive case, how often is it correct?

(TP)/Total Actual Positives

Classification Metrics

- Let's begin with recall.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Recall =
(TP)/5

- Recall:
  - When it actually is a positive case, how often is it correct?

(TP)/Total Actual Positives

Classification Metrics

- Let's begin with recall.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Recall =
(4)/5

- Recall:
  - When it actually is a positive case, how often is it correct?

(TP)/Total Actual Positives

Classification Metrics

- Let's begin with recall.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Recall = 0.8

- Recall:
  - How many relevant cases are found?

(TP)/Total Actual Positives

Classification Metrics

- What's the recall if we always classify as "healthy"?

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	0	0
	HEALTHY	5	95

Recall =
(TP)/Total Actual Positives

- Recall:
  - How many relevant cases are found?

(TP)/Total Actual Positives

Classification Metrics

- What's the recall if we always classify as "healthy"?

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	0	0
	HEALTHY	5	95

Recall =
(0)/5 !

- Recall:
  - How many relevant cases are found?

(TP)/Total Actual Positives

Classification Metrics

- A recall of 0 alerts you the model isn't catching cases!

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	0	0
	HEALTHY	5	95

Recall =
(0)/5 !

- Recall:
  - How many relevant cases are found?

(TP)/Total Actual Positives

Classification Metrics

- Now let's explore **precision**.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Precision =
(TP)/Total Predicted Positives

- Precision:
    - When prediction is positive, how often is it correct?

(TP)/Total Predicted Positives

Classification Metrics

- Now let's explore **precision**.

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Precision =
(TP)/6

- Precision:
    - When prediction is positive, how often is it correct?

(TP)/Total Predicted Positives

Classification Metrics

- Now let's explore **precision.**

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	4	2
	HEALTHY	1	93

Precision = 0.666

- Precision:
  - When prediction is positive, how often is it correct?

(TP)/Total Predicted Positives

Classification Metrics

- What's the **precision** if we always classify as "healthy"?

		ACTUAL	
		INFECTED	HEALTHY
PREDICTED	INFECTED	0	0
	HEALTHY	5	95

Precision =
(TP)/Total Predicted Positives

- Precision:
  - When prediction is positive, how often is it correct?

(TP)/Total Predicted Positives

وزارة الاتصــــــالات
وتكنولوجيا المعلومات

رواد مصر الرقمية

Classification Metrics

- Recall and Precision can help illuminate our performance specifically in regards to the relevant or positive case.
- Depending on the model, there is typically a trade-off between precision and recall, which we will explore later on with the ROC curve.

Classification Metrics

- Since precision and recall are related to each other through the numerator (TP), we often also report the F1-Score, which is the harmonic mean of precision and recall.

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

Classification Metrics

- The harmonic mean (instead of the normal mean) allows the entire harmonic mean to go to zero if **either** precision or recall ends up being zero.

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

Classification Metrics

- As a final note on the confusion matrix, there are **many** more metrics available:

		True condition				
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$	F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$		

Classification Metrics

- Finally, let’s explore a way to visualize the relationships between metrics such as precision and recall with curves.

Classification Performance Metrics

Part Three: ROC Curves

Classification Metrics

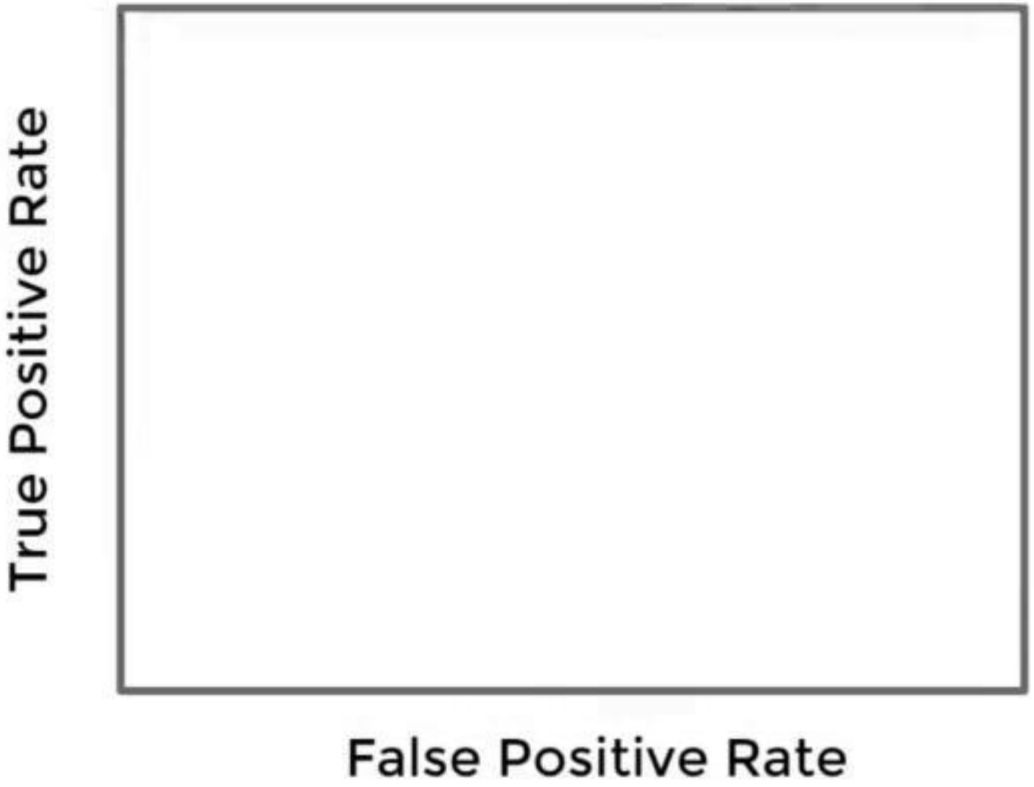
- During World War 2, Radar technology was developed to help detect incoming enemy aircraft.

Classification Metrics

- The technology was so new, the US Army wanted to develop a methodology to evaluate radar operator performance.

Classification Metrics

- They developed the Receiver Operator Characteristic curve.

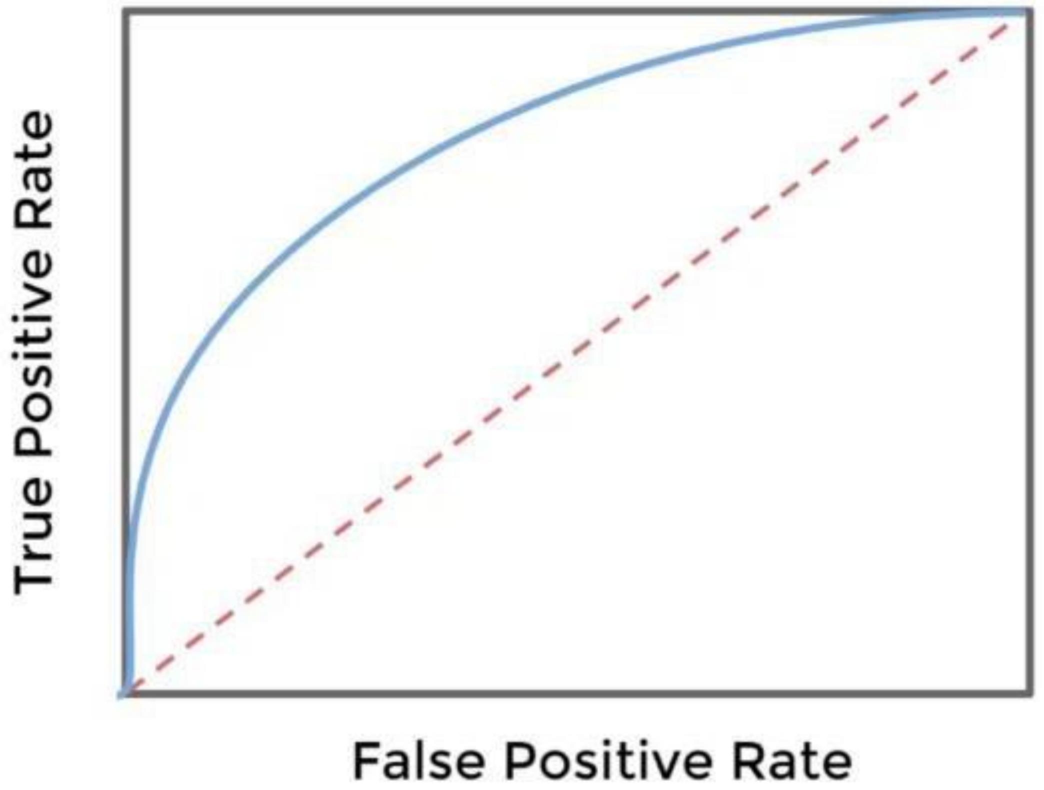


Classification Metrics

- They developed the Receiver Operator Characteristic curve.

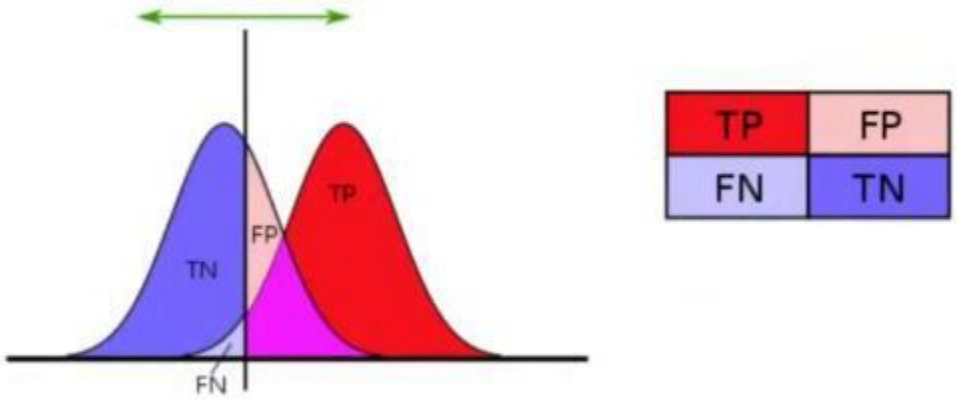
Classification Metrics

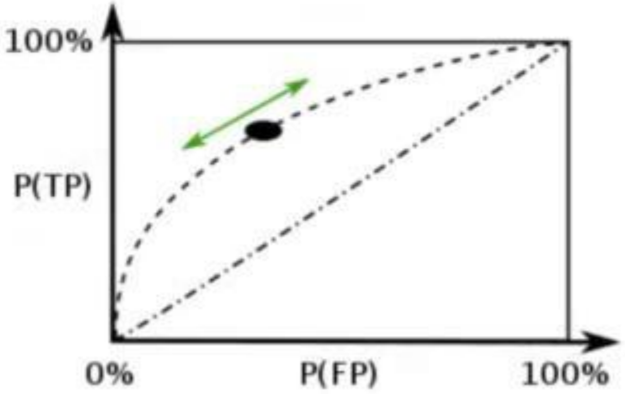
- There can be a trade-off between True Positives and False Positives.



Classification Metrics

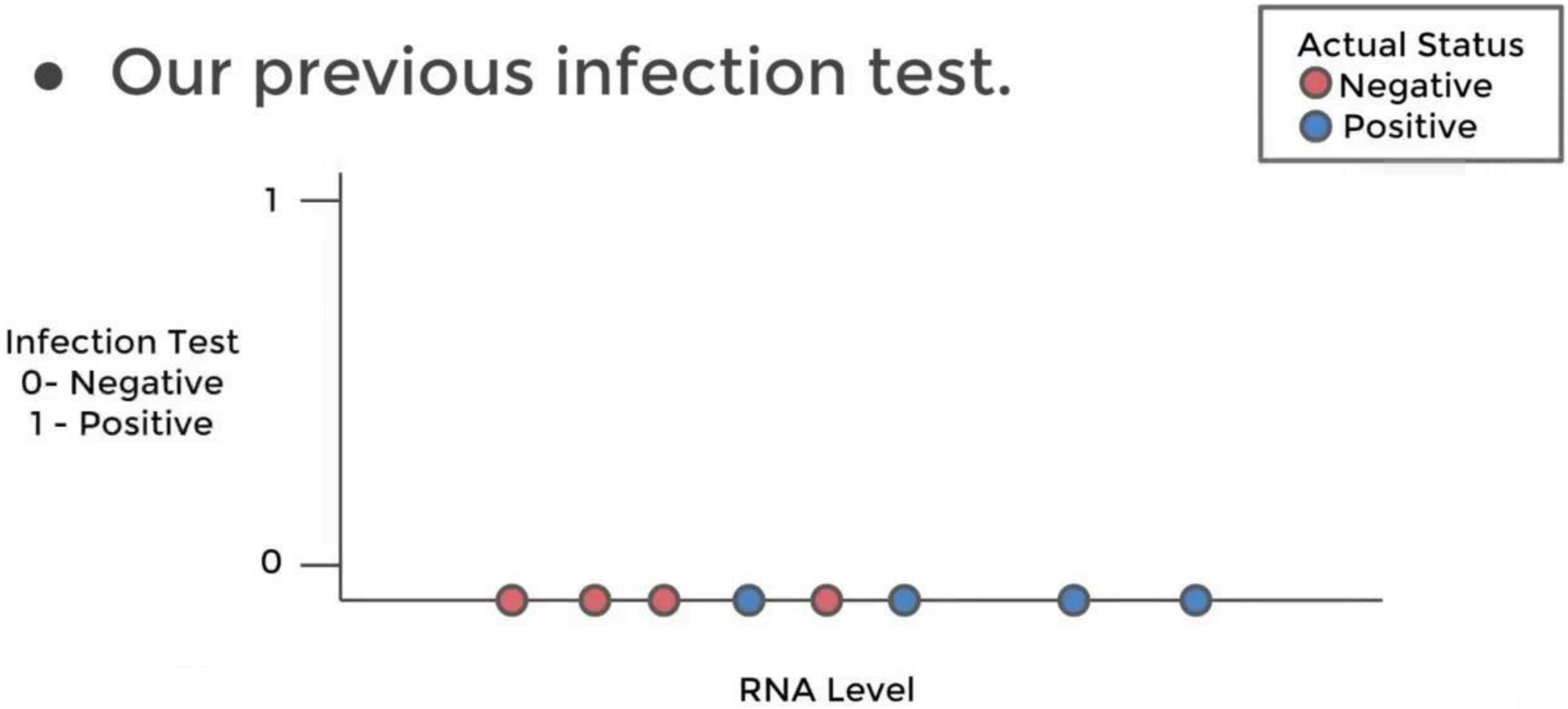
- There can be a trade-off between True Positives and False Positives.





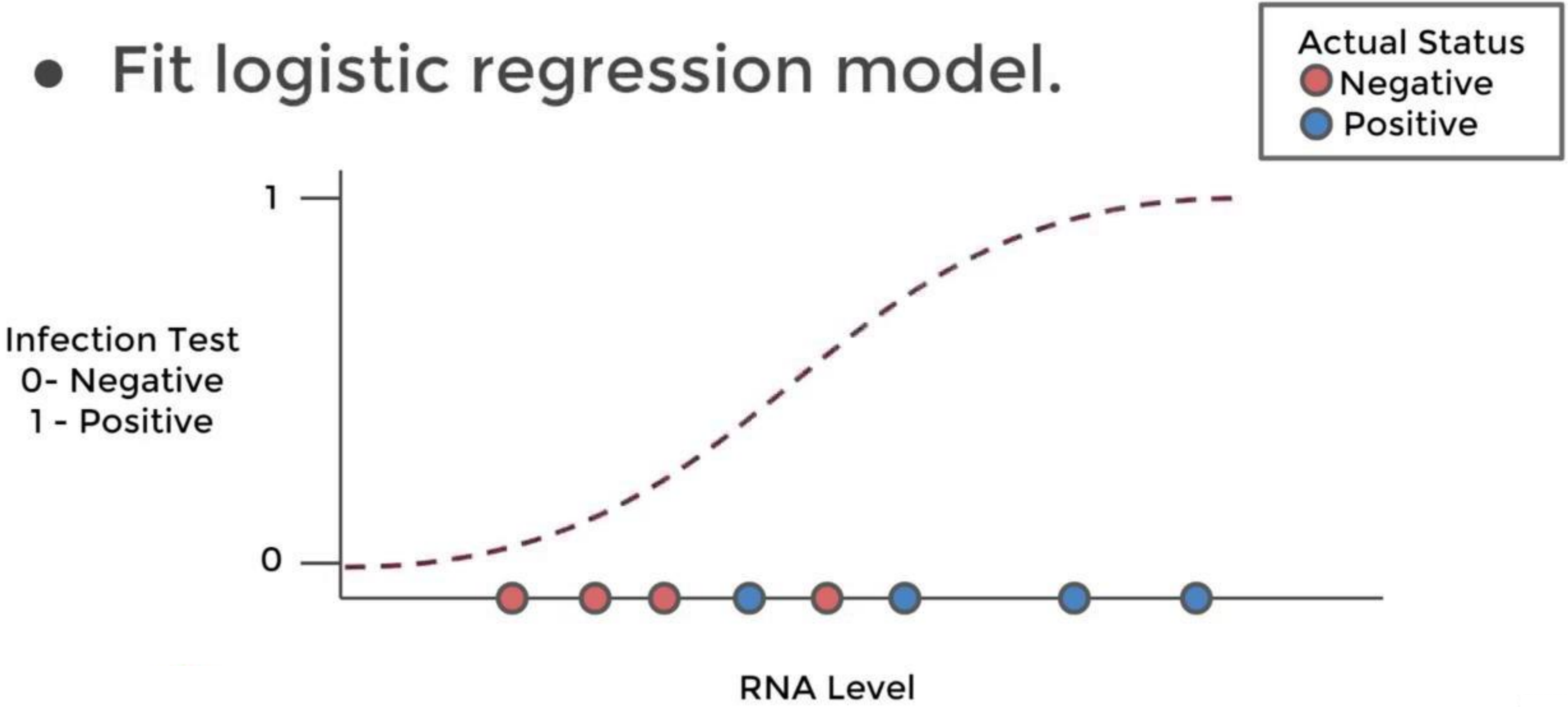
Classification Metrics

- Our previous infection test.



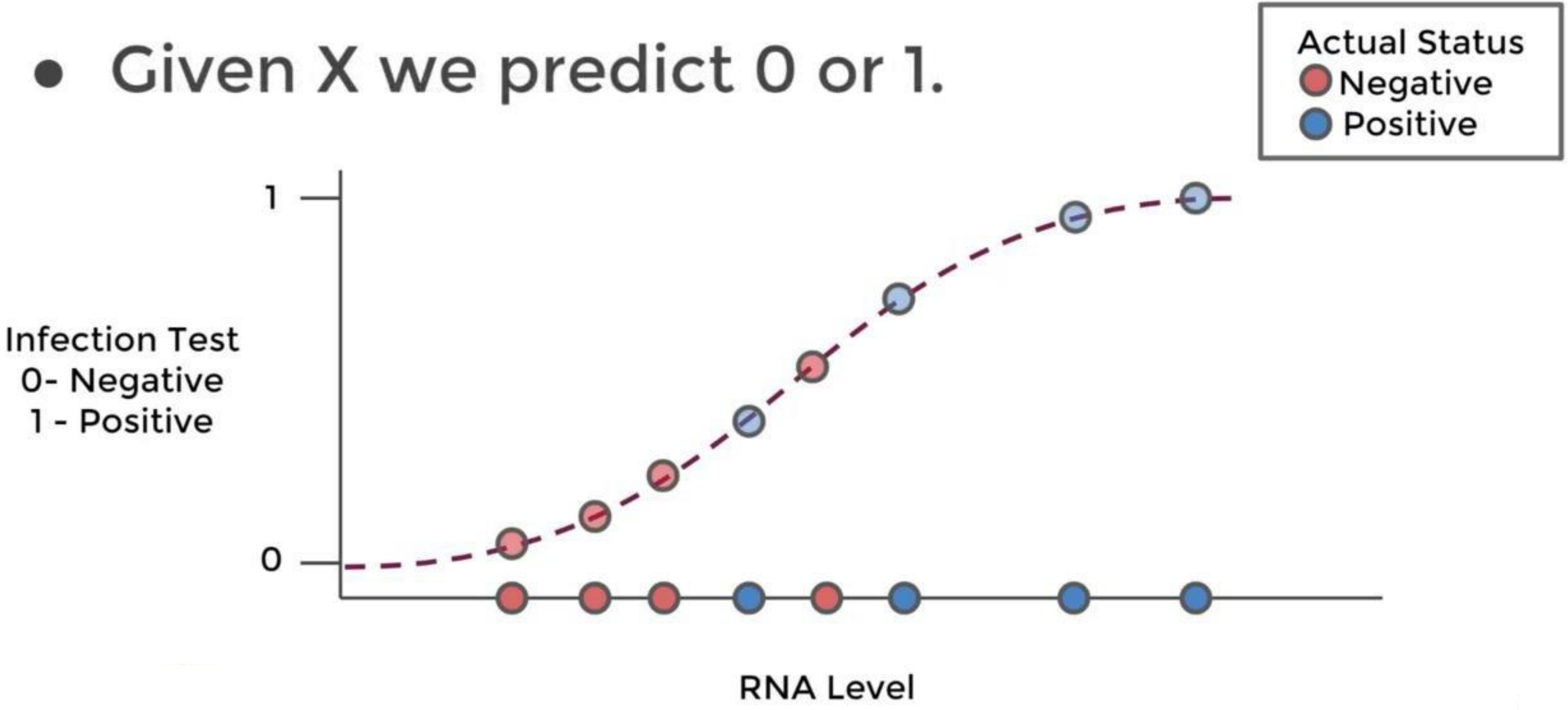
Classification Metrics

- Fit logistic regression model.



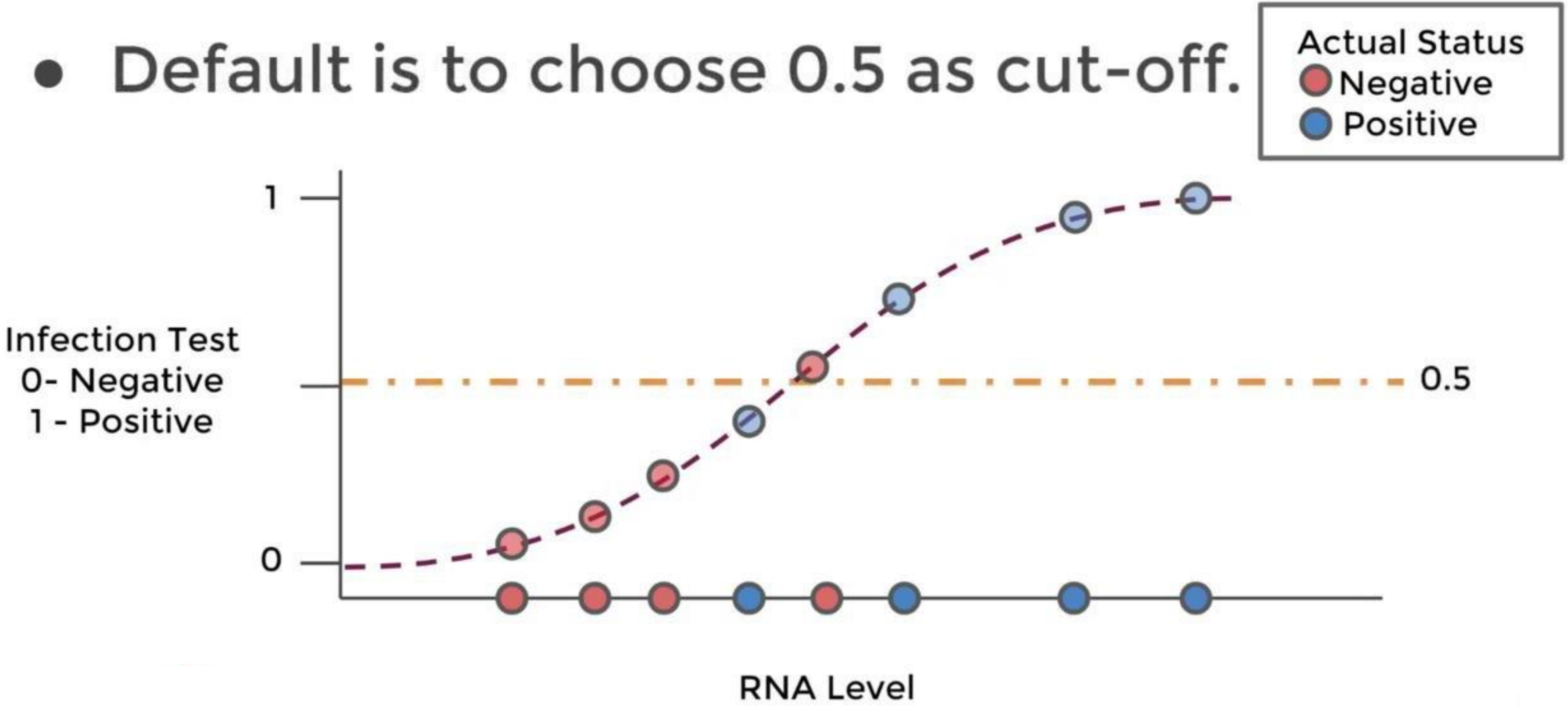
Classification Metrics

- Given X we predict 0 or 1.



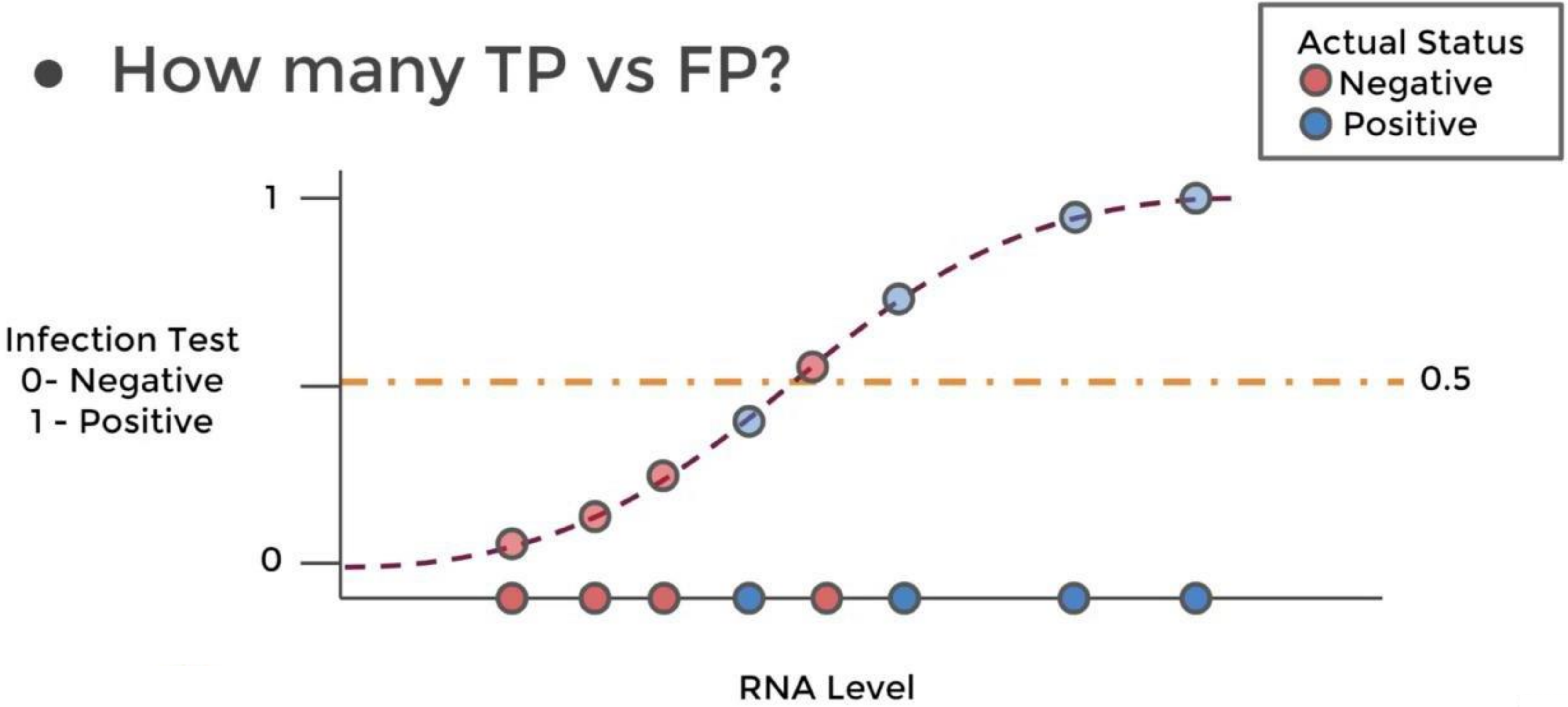
Classification Metrics

- Default is to choose 0.5 as cut-off.



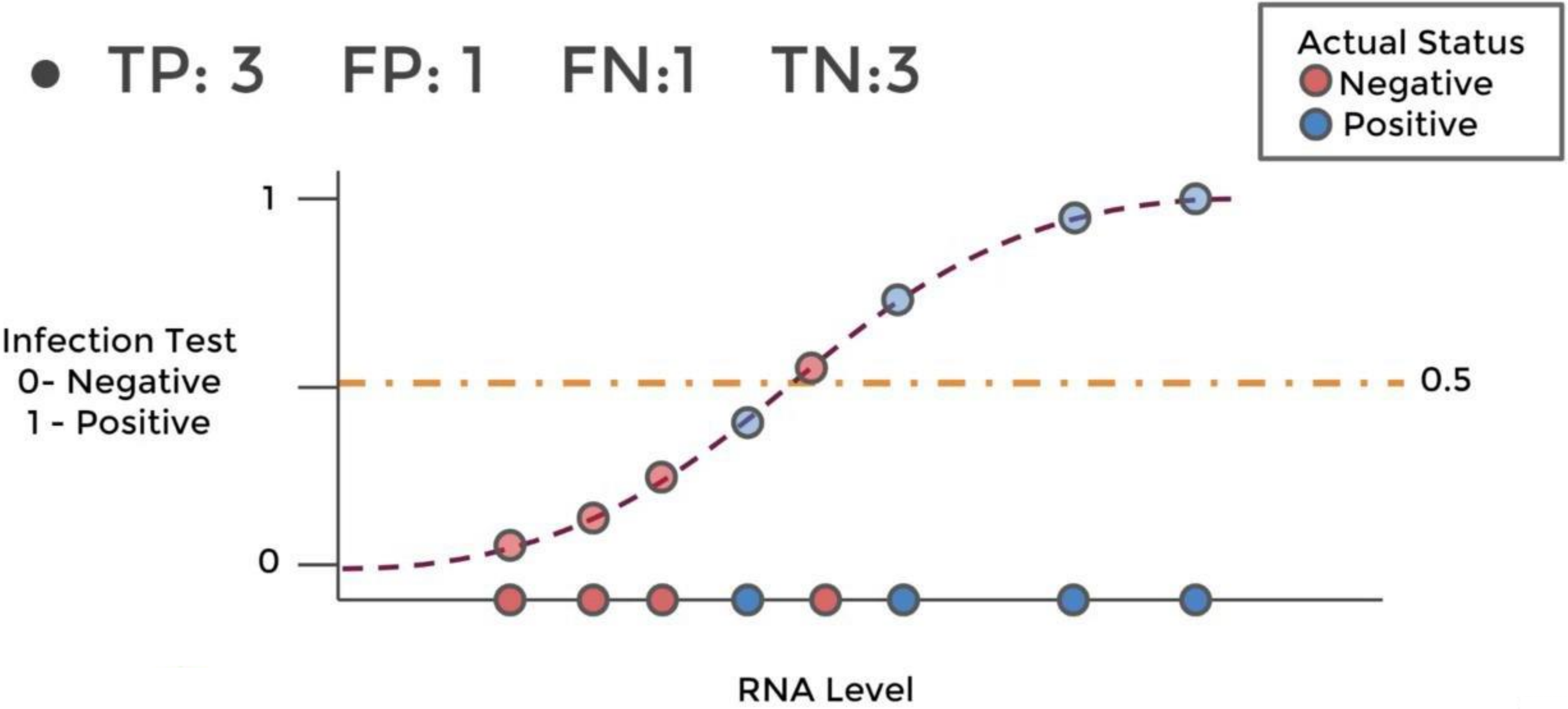
Classification Metrics

- How many TP vs FP?



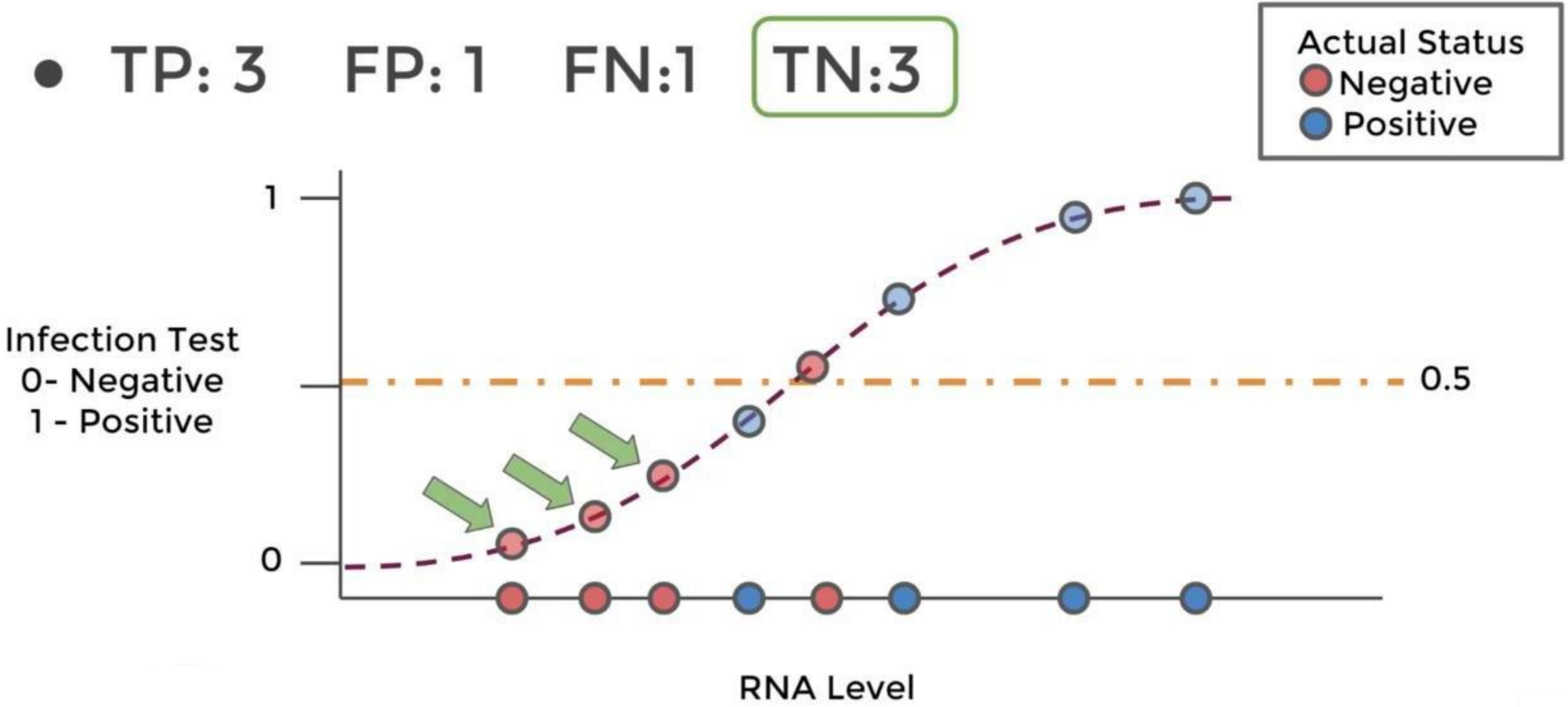
Classification Metrics

- TP: 3 FP: 1 FN:1 TN:3



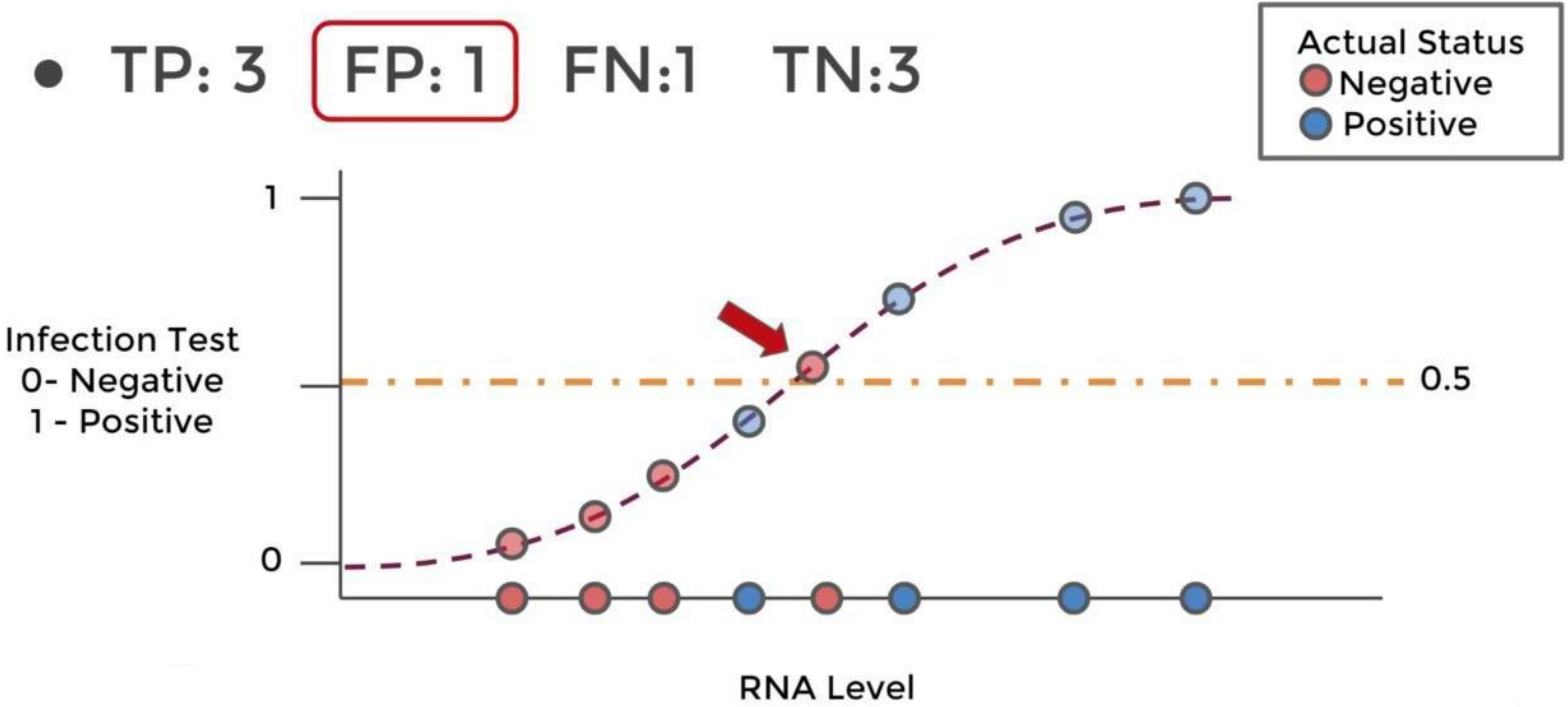
Classification Metrics

- TP: 3 FP: 1 FN:1 $\boxed{\text{TN:3}}$



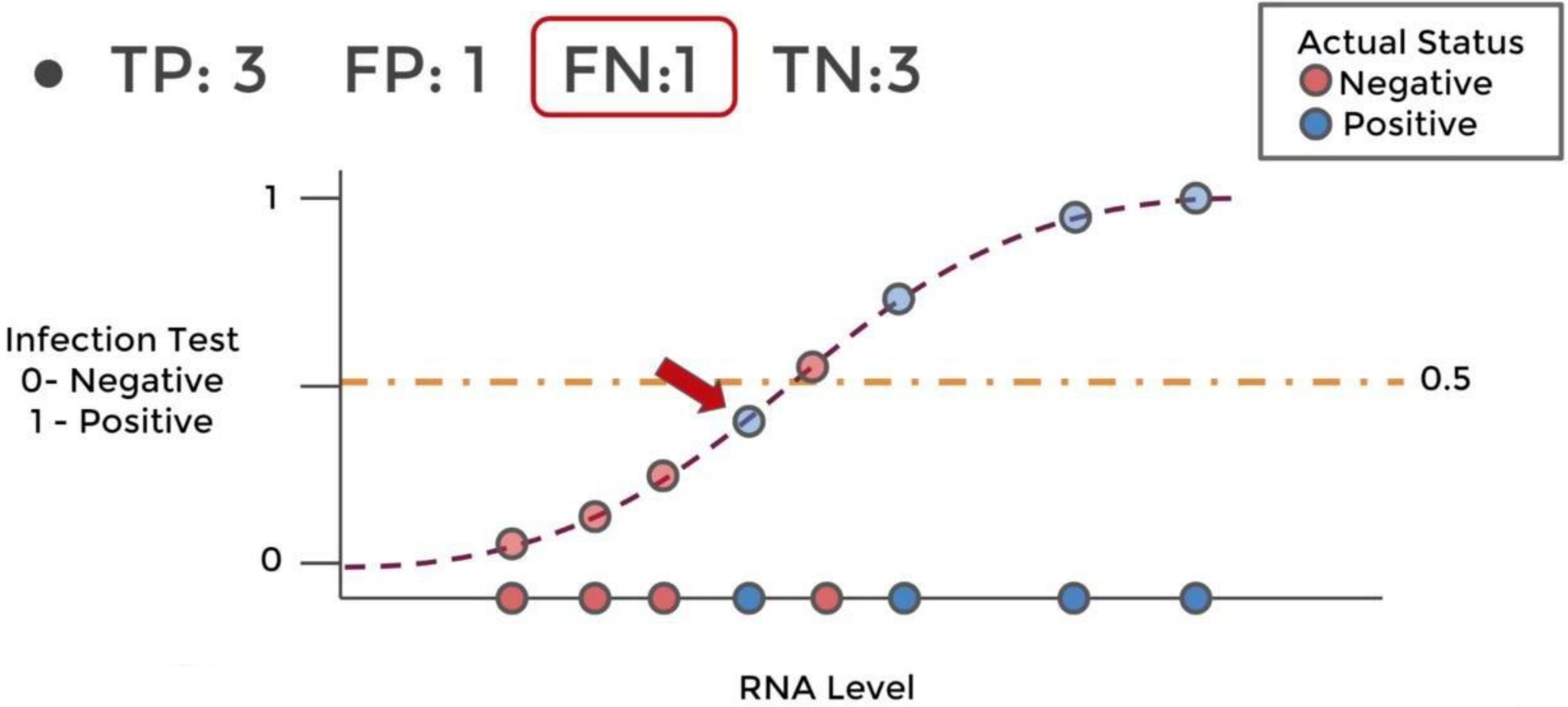
Classification Metrics

- TP: 3 FP: 1 FN:1 TN:3



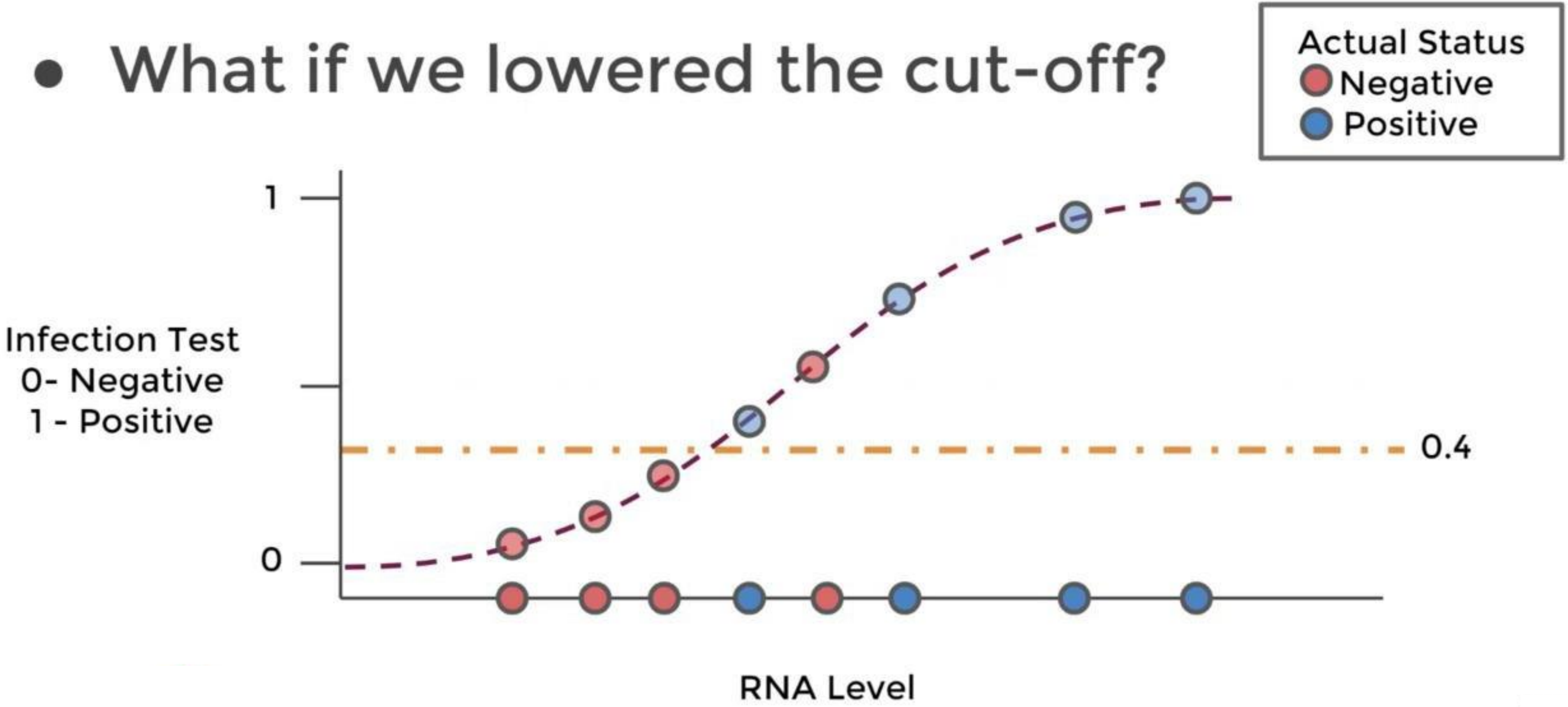
Classification Metrics

- TP: 3 FP: 1 $\boxed{\text{FN:1}}$ TN:3



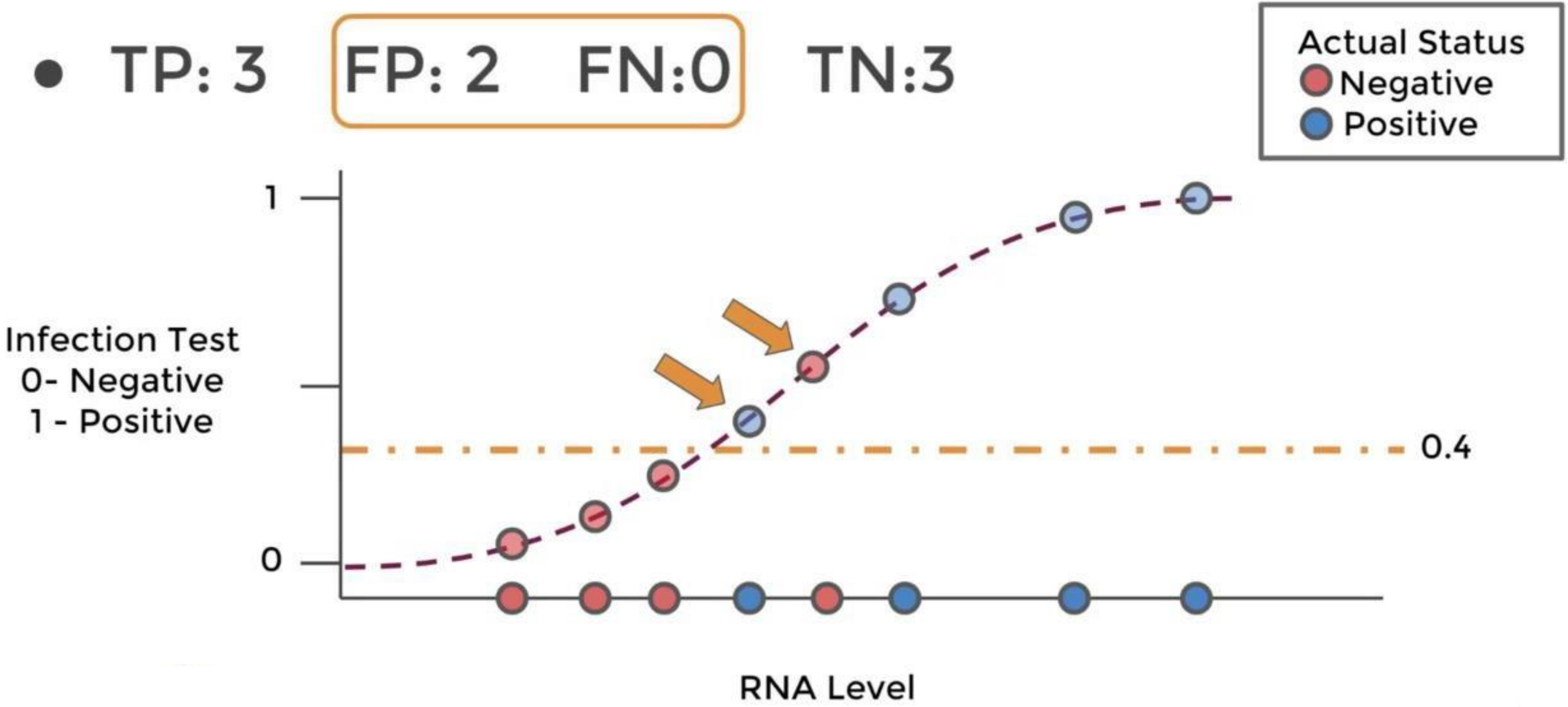
Classification Metrics

- What if we lowered the cut-off?



Classification Metrics

- TP: 3 FP: 2 FN:0 TN:3

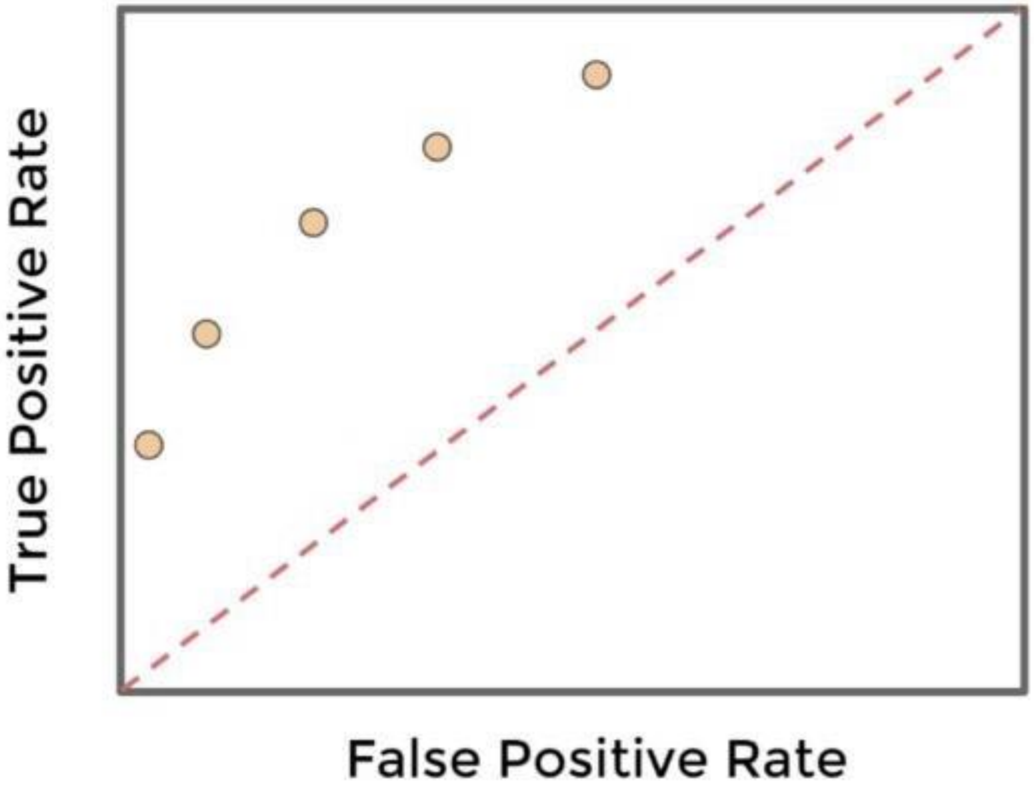


Classification Metrics

- In certain situations, we gladly accept more false positives to reduce false negatives.
- Imagine a dangerous virus test, we would much rather produce false positives and later do more stringent examination than accidentally release a false negative!

Classification Metrics

- Chart the True vs. False positives for various cut-offs for the ROC curve.

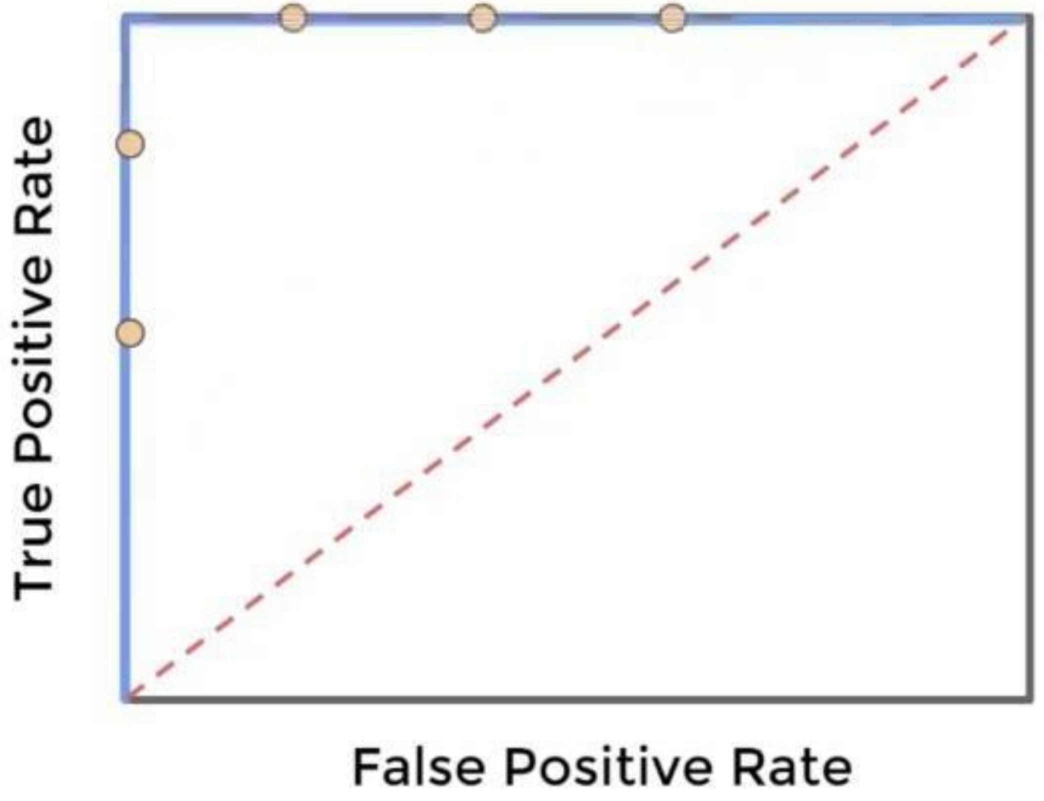


Classification Metrics

- By changing the cut-off limit, we can adjust our True vs. False Positives!

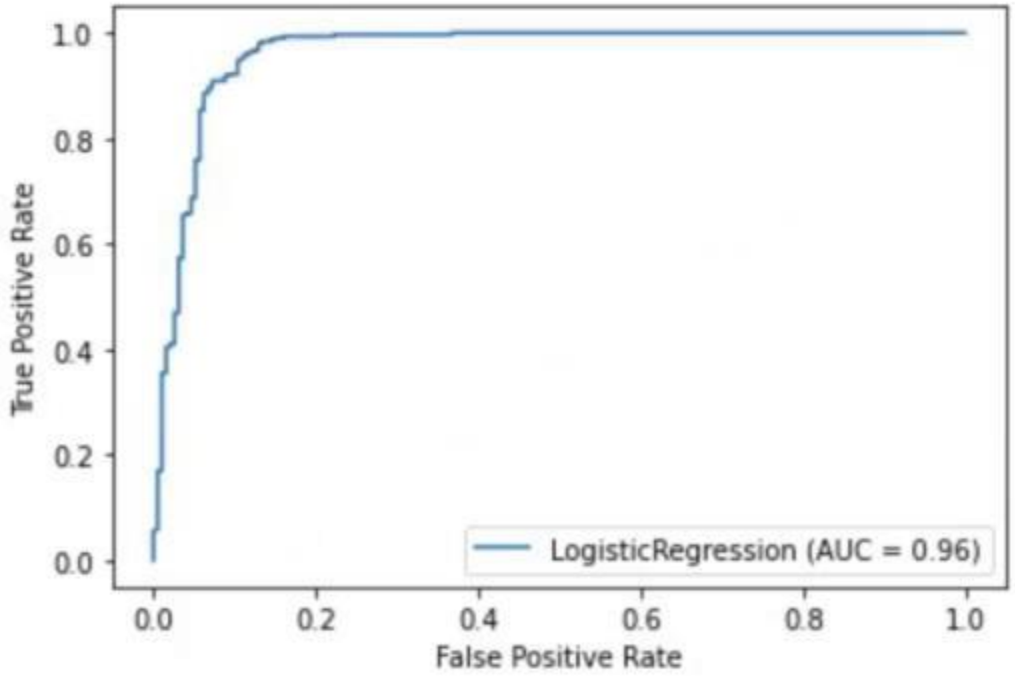
Classification Metrics

- A perfect model would have a zero FPR.
- Random guessing is the red line.



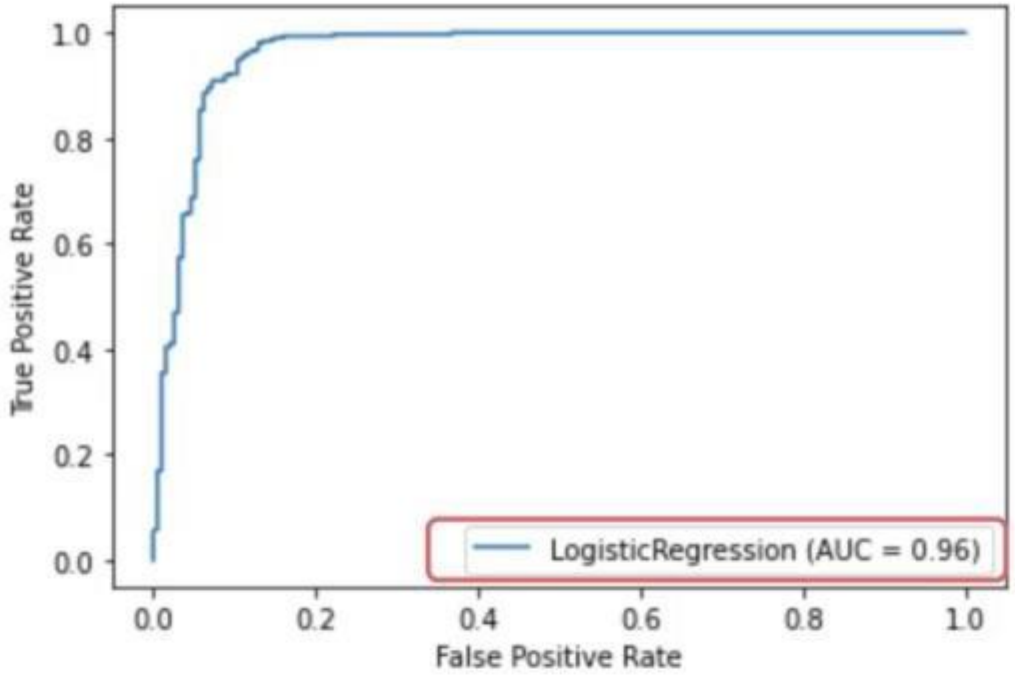
Classification Metrics

- Realistically with smaller data sets the ROC curves are not as smooth.



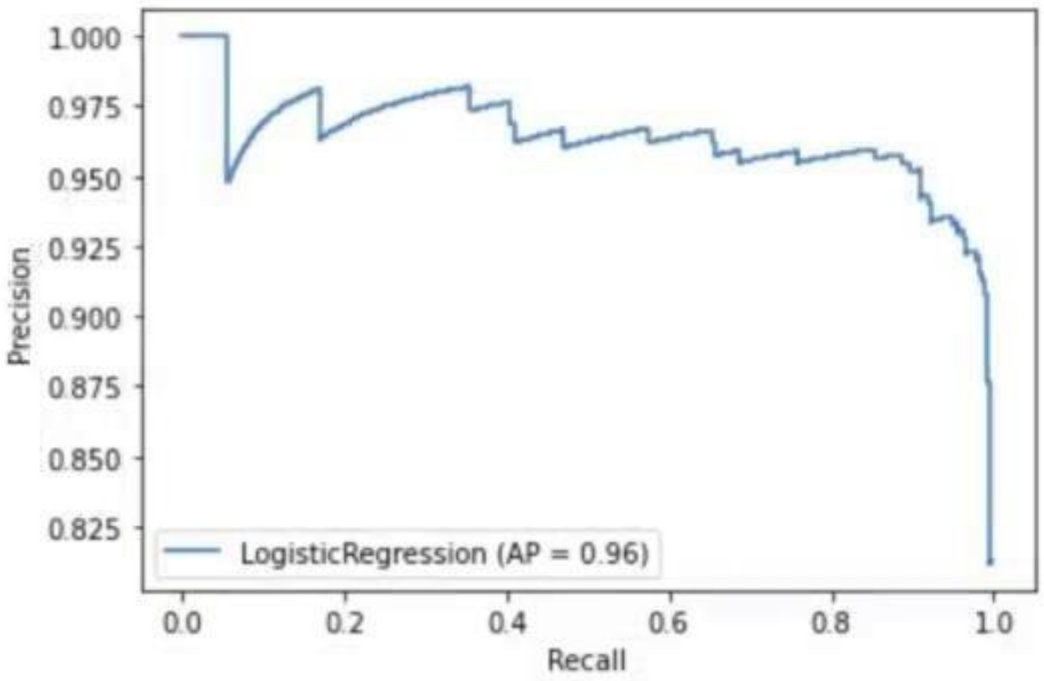
Classification Metrics

- AUC - Area Under the Curve , allows us to compare ROCs for different models.



Classification Metrics

- Can also create precision vs. recall curves:



Logistic Regression with Scikit-Learn

Part Three: Performance Metrics

00-Logistic-Regression.ipynb