

# Machine Learning I

Mohamed Hussien

# Classification Model Performance

# Lecture Overview

**Accuracy**

**False positives & False negatives**

**Confusion Matrix**

**Recall & Precision**

**F-score**

**Sensitivity and Specificity**

**ROC (Receiver Operating Characteristic)**

**AUC (Area Under the Curve)**

**Decision based on ROC**

# Accuracy

“

The ratio between the number of correctly predicted data points and the total number of data points.

# Accuracy

“

The ratio between the number of correctly predicted data points and the total number of data points.

If we evaluate a model on a test dataset of 1000 samples, and the model predicted the correct label of the sample 875 times, then this model has an accuracy of  $875/1000 = 0.875$ , or 87.5%.

If we have 3 models for binary problem with 15%, 50% and 70% accuracy. Order them from the best?

Is it always enough for model evaluation?

# Accuracy



# Accuracy

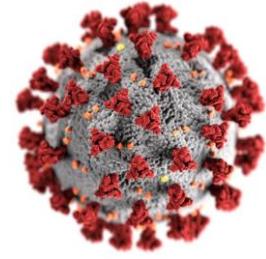
We will use 2 datasets in this lecture:

- Medical Dataset - A set of patient diagnosed with coronavirus  
A medical dataset with 1000 persons.  
**10 diagnosed with coronavirus**  
**990 diagnosed as healthy**  
Labels “sick” or “healthy”.

# Accuracy

We will use 2 datasets in this lecture:

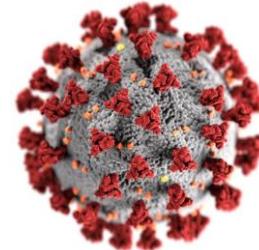
- Medical Dataset - A set of patient diagnosed with coronavirus  
A medical dataset with 1000 persons.  
**10 diagnosed with coronavirus**  
**990 diagnosed as healthy**  
Labels “sick” or “healthy”.



# Accuracy

We will use 2 datasets in this lecture:

- Medical Dataset - A set of patient diagnosed with coronavirus  
A medical dataset with 1000 persons.  
**10 diagnosed with coronavirus**  
**990 diagnosed as healthy**  
Labels “sick” or “healthy”.
- Email Dataset - A set of emails labeled spam or ham  
A dataset of 100 emails.  
**40 are spam**  
**60 are ham**  
Labels are “spam” and “ham”



# Accuracy



# Accuracy

Let's back to our question "Is accuracy always enough for model evaluation?"

Suppose I tell you the following: "I have developed a test for coronavirus that takes 10 seconds to run, doesn't require any examinations, and has an accuracy of 99%!".

What do you think?

On our coronavirus dataset, If I tested all samples as healthy, my model accuracy is 99% !!!

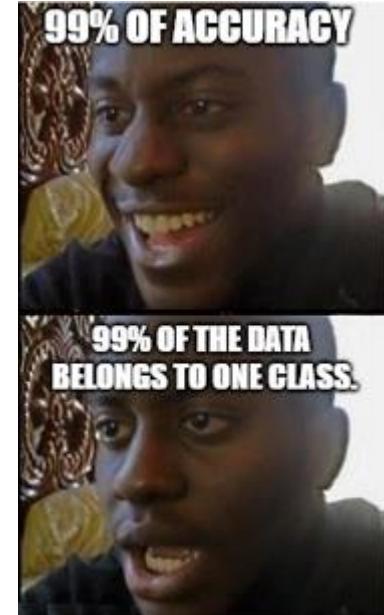
# Accuracy

Let's back to our question "Is accuracy always enough for model evaluation?"

Suppose I tell you the following: "I have developed a test for coronavirus that takes 10 seconds to run, doesn't require any examinations, and has an accuracy of 99%!".

What do you think?

On our coronavirus dataset, If I tested all samples as healthy, my model accuracy is 99% !!!



# Accuracy

Let's back to our question "Is accuracy always enough for model evaluation?"

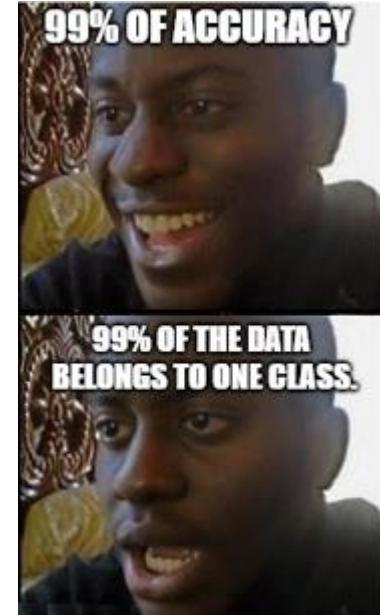
Suppose I tell you the following: "I have developed a test for coronavirus that takes 10 seconds to run, doesn't require any examinations, and has an accuracy of 99%!".

What do you think?

On our coronavirus dataset, If I tested all samples as healthy, my model accuracy is 99% !!!

What is the problem of our model?

The problem is that errors are not created equal, and some mistakes are much more expensive than others.



# Accuracy



Use colab to open this github notebook:

[“s7s/machine\\_learning\\_1/classification\\_model\\_performance/classification\\_model\\_performance.ipynb”](https://colab.research.google.com/github/s7s/machine_learning_1/blob/main/classification_model_performance/classification_model_performance.ipynb)

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

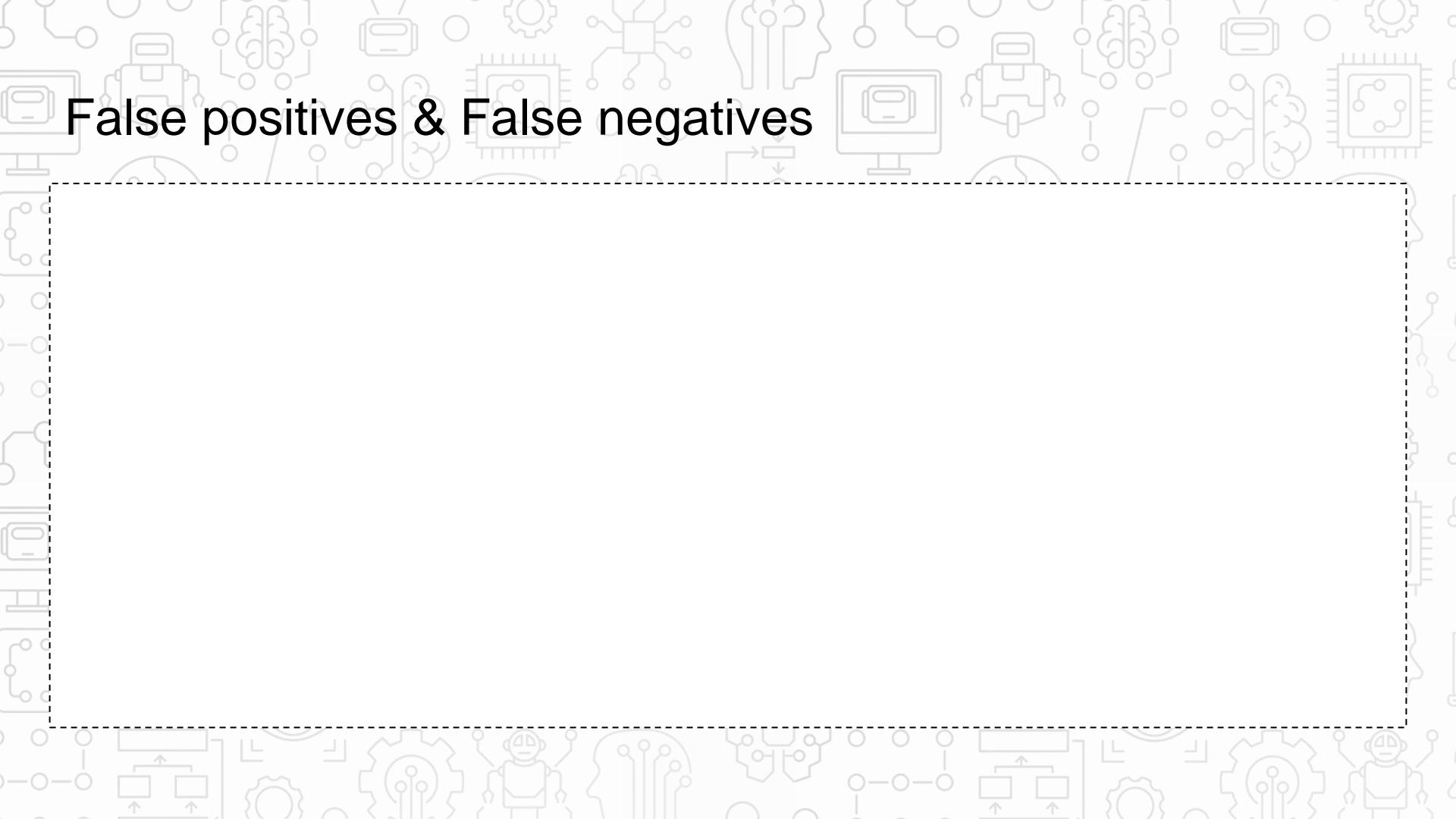
Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

Decision based on ROC

# False positives & False negatives



# False positives & False negatives

For coronavirus dataset:

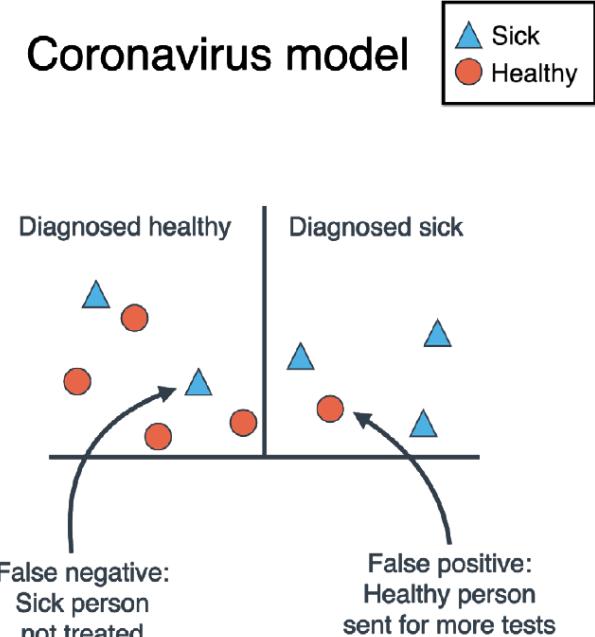
- False positive: A healthy person that gets incorrectly diagnosed as sick.
- False negative: A sick person that gets incorrectly diagnosed as healthy.
- True positive: A sick person that gets diagnosed as sick.
- True negative: A healthy person that gets diagnosed as healthy.

# False positives & False negatives

For coronavirus dataset:

- False positive: A healthy person that gets incorrectly diagnosed as sick.
- False negative: A sick person that gets incorrectly diagnosed as healthy.
- True positive: A sick person that gets diagnosed as sick.
- True negative: A healthy person that gets diagnosed as healthy.

Coronavirus model



# False positives & False negatives

For coronavirus dataset:

- False positive: A healthy person that gets incorrectly diagnosed as sick.
- False negative: A sick person that gets incorrectly diagnosed as healthy.
- True positive: A sick person that gets diagnosed as sick.
- True negative: A healthy person that gets diagnosed as healthy.

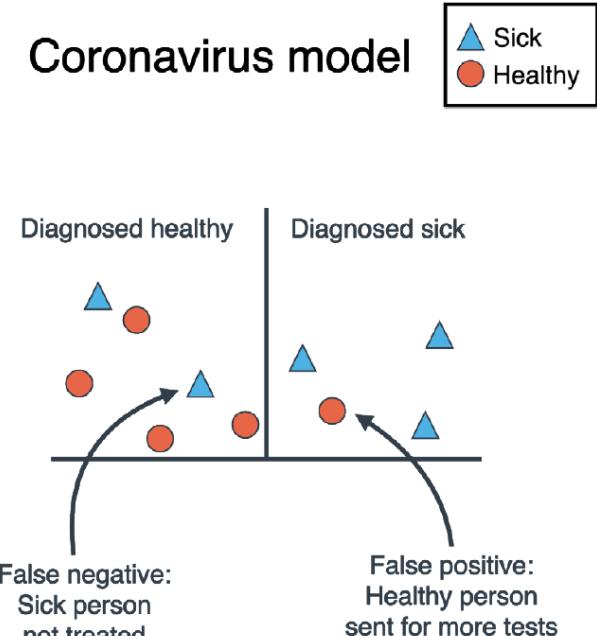
3 true positives

4 true negatives

1 false positive

2 false negatives

Coronavirus model



# False positives & False negatives

For coronavirus dataset:

- False positive: A healthy person that gets incorrectly diagnosed as sick.
- False negative: A sick person that gets incorrectly diagnosed as healthy.
- True positive: A sick person that gets diagnosed as sick.
- True negative: A healthy person that gets diagnosed as healthy.

3 true positives

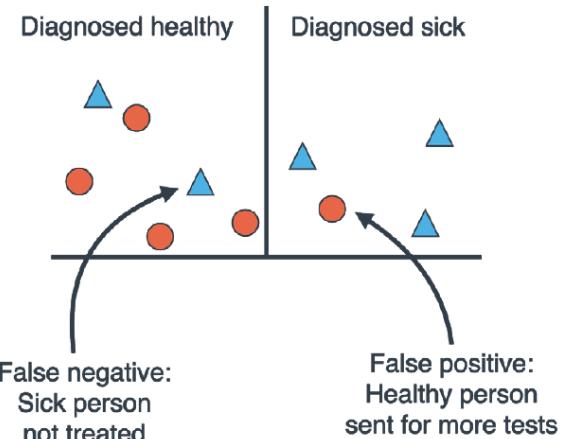
4 true negatives

1 false positive

2 false negatives

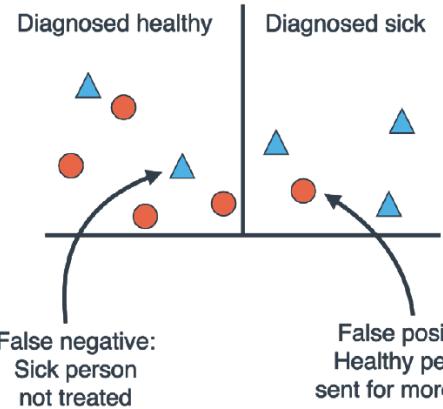
Find the same for the Spam dataset..

Coronavirus model

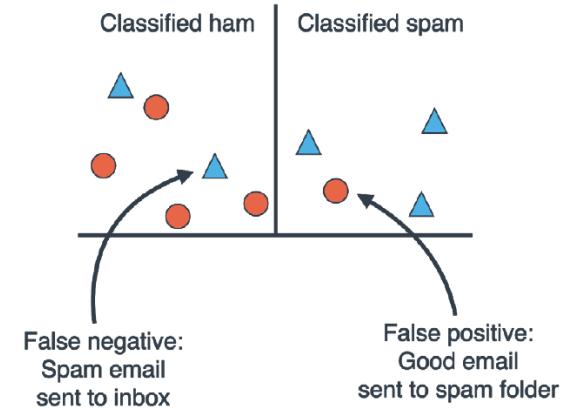
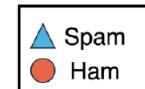


# False positives & False negatives

Coronavirus model



Spam model



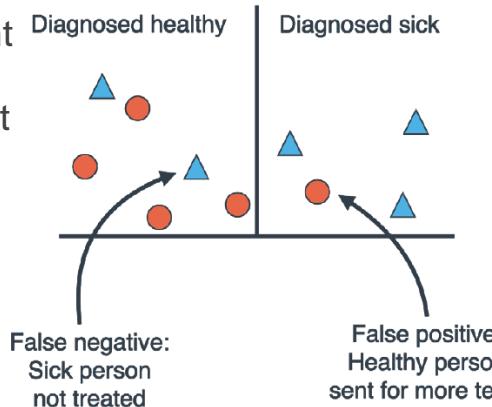
# False positives & False negatives

Which is important for each dataset  
False positives or False negatives?

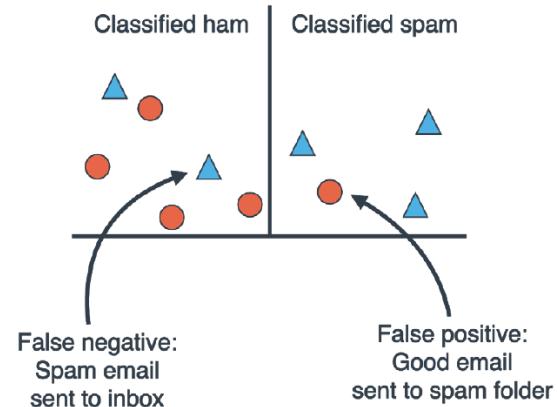
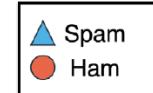
In coronavirus it is more important  
to not have undetected sick people  
So false negative is more important

In spam model it is more important  
to not have ham mails in junk box  
So false positive is more important

Coronavirus model



Spam model



# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

Decision based on ROC

# Confusion Matrix



# Confusion Matrix

It is called that because it tells how much the model is confused between 2 classes.

Number of rows = Number of columns = Number of classes.

In the rows we write the true labels (in the medical example, this is the condition of the person, sick or healthy) and in the columns we write the predicted labels (the diagnosis of the person, sick or healthy).

# Confusion Matrix

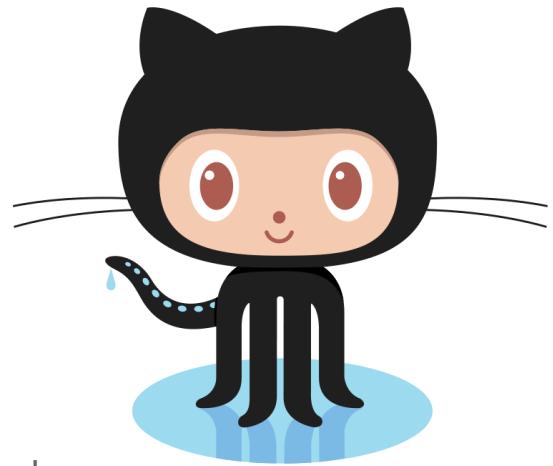
It is called that because it tells how much the model is confused between 2 classes.

Number of rows = Number of columns = Number of classes.

In the rows we write the true labels (in the medical example, this is the condition of the person, sick or healthy) and in the columns we write the predicted labels (the diagnosis of the person, sick or healthy).

Person's condition	Predicted positive	Predicted negative
Positive	Number of true positives	Number of false negatives
Negative	Number of false positives	Number of true negatives

# Confusion Matrix



Use colab to open this github notebook:

[“s7s/machine\\_learning\\_1/classification\\_model\\_performance/classification\\_model\\_performance.ipynb”](https://colab.research.google.com/github/s7s/machine_learning_1/blob/main/classification_model_performance/classification_model_performance.ipynb)

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

Decision based on ROC

# Recall



# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

Coronavirus Model 1	Diagnosed sick (predicted positive)	Diagnosed healthy (predicted negative)
Sick (positive)	(number of true positives)	(number of false negatives)
Healthy (negative)	(number of false positives)	(number of true negatives)

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

Coronavirus Model 1	Diagnosed sick (predicted positive)	Diagnosed healthy (predicted negative)
Sick (positive)	0 (number of true positives)	(number of false negatives)
Healthy (negative)	(number of false positives)	(number of true negatives)

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

Coronavirus Model 1	Diagnosed sick (predicted positive)	Diagnosed healthy (predicted negative)
Sick (positive)	0 (number of true positives)	(number of false negatives)
Healthy (negative)	0 (number of false positives)	(number of true negatives)

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

Coronavirus Model 1	Diagnosed sick (predicted positive)	Diagnosed healthy (predicted negative)
Sick (positive)	0 (number of true positives)	10 (number of false negatives)
Healthy (negative)	0 (number of false positives)	(number of true negatives)

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

Coronavirus Model 1	Diagnosed sick (predicted positive)	Diagnosed healthy (predicted negative)
Sick (positive)	0 (number of true positives)	10 (number of false negatives)
Healthy (negative)	0 (number of false positives)	990 (number of true negatives)

# Recall

It finds the proportion of correct positive predictions among the data points with a positive **label**.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

Coronavirus Model 1	Diagnosed sick (predicted positive)	Diagnosed healthy (predicted negative)
Sick (positive)	0 (number of true positives)	10 (number of false negatives)
Healthy (negative)	0 (number of false positives)	990 (number of true negatives)

Coronavirus Model has a total of 0 true positives among 10 positive-labeled, so its recall is  $0/10 = 0$ .

# Precision



# Precision

It finds the proportion of correct positive predictions among the data points with a positive **prediction**.

# Precision

It finds the proportion of correct positive predictions among the data points with a positive **prediction**.

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

# Precision

It finds the proportion of correct positive predictions among the data points with a positive **prediction**.

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

Spam Model 1	Predicted spam	Predicted ham
Spam	30 (true positives)	10 (false negatives)
Ham	5 (false positives)	55 (true negatives)

# Precision

It finds the proportion of correct positive predictions among the data points with a positive **prediction**.

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

Spam Model 1	Predicted spam	Predicted ham
Spam	30 (true positives)	10 (false negatives)
Ham	5 (false positives)	55 (true negatives)

Spam Model has a total of 30 true positives among 35 positive-predicted, so its precision is  $30/35 = 86.7\%$

# Recall VS Precision

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}.$$

# Recall VS Precision

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

# Recall VS Precision

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

Positive labels that detected as negative  
Ex: corona patients diagnosed healthy

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

# Recall VS Precision

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

Positive labels that detected as negative  
Ex: corona patients diagnosed healthy

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

Negative labels that detected as positive  
Ex: Ham mails detected as spam

# Recall VS Precision

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

Positive labels that detected as negative  
Ex: corona patients diagnosed healthy

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

Negative labels that detected as positive  
Ex: Ham mails detected as spam

In the coronavirus model, False negatives is more important to be decreased as positive people should be diagnosed as positive, otherwise the pandemic will spread. **High recall needed.**

In the spam model, False positives is more important to be decreased as Ham mails should be detected as Ham, otherwise They will lost in junk. **High precision needed.**

We need a metric that combines both metrics with weights.

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

Decision based on ROC

# F-score



# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

$a=1$  and  $b=9$ , their average is 5.

The harmonic mean is 1.8 which is better for our case.

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

$$F_{\beta} = \frac{(1+\beta^2) P R}{\beta^2 P + R}.$$

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

If  $\beta = 1$

$$F_1 = \frac{2PR}{P + R}.$$

Recall and precision has the same weight

$$F_\beta = \frac{(1+\beta^2) PR}{\beta^2 P + R}.$$

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

If  $\beta = 1$

$$F_1 = \frac{2PR}{P + R}.$$

If  $\beta > 1$  say 10

$$\frac{101 P R}{100P + R}.$$

Recall and precision has the same weight

Recall has a weight more than precision

$$F_\beta = \frac{(1+\beta^2) P R}{\beta^2 P + R}.$$

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

If  $\beta = 1$

$$F_1 = \frac{2PR}{P + R}$$

Recall and precision has the same weight

If  $\beta > 1$  say 10

$$\frac{101 P R}{100P + R}$$

Recall has a weight more than precision

If  $\beta < 1$  say 0.1

$$\frac{1.01 P R}{0.01P + R}$$

Precision has a weight more than recall

$$F_\beta = \frac{(1+\beta^2) P R}{\beta^2 P + R}$$

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

R= 0.1

If  $\beta = 1$

P= 0.9

$$F_1 = \frac{2PR}{P + R}$$

Recall and precision has the same weight

If  $\beta > 1$  say 10

$$\frac{101 P R}{100P + R}$$

Recall has a weight more than precision

If  $\beta < 1$  say 0.1

$$\frac{1.01 P R}{0.01P + R}$$

Precision has a weight more than recall

$$F_\beta = \frac{(1+\beta^2) P R}{\beta^2 P + R}$$

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

R= 0.1

If  $\beta = 1$

P= 0.9

$$F_1 = \frac{2PR}{P + R} = 0.18$$

If  $\beta > 1$  say 10

$$\frac{101 P R}{100P + R}$$

If  $\beta < 1$  say 0.1

$$\frac{1.01 P R}{0.01P + R}$$

Recall and precision has the same weight

Recall has a weight more than precision

Precision has a weight more than recall

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

R= 0.1

If  $\beta = 1$

P= 0.9

$$F_1 = \frac{2PR}{P + R} = 0.18$$

If  $\beta > 1$  say 10

$$\frac{101 P R}{100P + R} = 0.10$$

If  $\beta < 1$  say 0.1

$$\frac{1.01 P R}{0.01P + R}$$

Recall and precision has the same weight

Recall has a weight more than precision

Precision has a weight more than recall

# F-score

How to combine between the recall and precision in one metric?

Average is not the best option.

If a model has, say, recall of 50% and precision of 100%, the average is 75%. This is a good score, but the model may not be, since a recall of 50% is not very good.

We use **harmonic mean**:  $2ab/(a+b)$

a=1 and b=9, their average is 5.

The harmonic mean is 1.8 which is better for our case.

R= 0.1

If  $\beta = 1$

P= 0.9

$$F_1 = \frac{2PR}{P + R} = 0.18$$

If  $\beta > 1$  say 10

$$\frac{101 P R}{100P + R} = 0.10$$

If  $\beta < 1$  say 0.1

$$\frac{1.01 P R}{0.01P + R} = 0.83$$

Recall and precision has the same weight

Recall has a weight more than precision

Precision has a weight more than recall

# Recall, Precision, and F-Score



Use colab to open this github notebook:

[“s7s/machine\\_learning\\_1/classification\\_model\\_performance/classification\\_model\\_performance.ipynb”](https://colab.research.google.com/github/s7s/machine_learning_1/blob/main/classification_model_performance/classification_model_performance.ipynb)

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

Decision based on ROC

# Sensitivity and Specificity



# Sensitivity and Specificity

**Sensitivity (true positive rate):** It finds the proportion of **correct positive** predictions among the data points with **positive label**. **[Recall]**

# Sensitivity and Specificity

**Sensitivity (true positive rate):** It finds the proportion of **correct positive** predictions among the data points with **positive label**. [Recall]

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

# Sensitivity and Specificity

**Sensitivity (true positive rate):** It finds the proportion of **correct positive** predictions among the data points with **positive label**. **[Recall]**

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

**Specificity (true negative rate):** It finds the proportion of **correct Negative** predictions among the data points with **Negative label**.

# Sensitivity and Specificity

**Sensitivity (true positive rate):** It finds the proportion of **correct positive** predictions among the data points with **positive label**. **[Recall]**

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

**Specificity (true negative rate):** It finds the proportion of **correct Negative** predictions among the data points with **Negative label**.

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}}.$$

# Sensitivity and Specificity

**Sensitivity (true positive rate):** It finds the proportion of **correct positive** predictions among the data points with **positive label**. **[Recall]**

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}.$$

**Specificity (true negative rate):** It finds the proportion of **correct Negative** predictions among the data points with **Negative label**.

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}}.$$

In the coronavirus model, the sensitivity is the proportion of sick people that the model has correctly diagnosed, among all the sick people. The specificity is the proportion of healthy people the model has correctly diagnosed, among the healthy people. Since we are more concerned about correctly diagnosing the sick people, we need the coronavirus model to have **high sensitivity**.

# Sensitivity and Specificity

**Sensitivity (true positive rate):** It finds the proportion of **correct positive** predictions among the data points with **positive label**. **[Recall]**

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

**Specificity (true negative rate):** It finds the proportion of **correct Negative** predictions among the data points with **Negative label**.

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}}$$

In the coronavirus model, the sensitivity is the proportion of sick people that the model has correctly diagnosed, among all the sick people. The specificity is the proportion of healthy people the model has correctly diagnosed, among the healthy people. Since we are more concerned about correctly diagnosing the sick people, we need the coronavirus model to have **high sensitivity**.

In the spam detection model, the sensitivity is the proportion of spam messages we correctly deleted, among all the spam messages. The specificity is the proportion of ham emails we correctly sent to the inbox, among all the ham emails. Since we are more concerned about correctly detecting the ham emails, we need the spam detection model to have **high specificity**.

# Recall , Precision, Sensitivity and Specificity

	Predicted positive	Predicted negative
Positive label	True positives False negatives	
Negative label	False positives	True negatives

$$\text{recall (sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

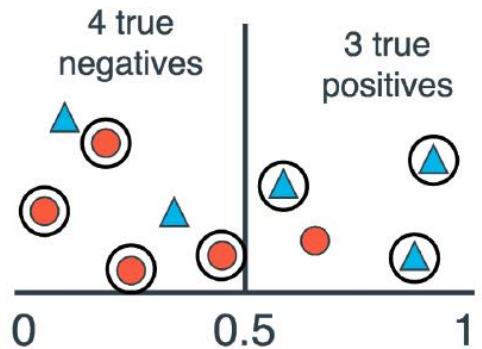
Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

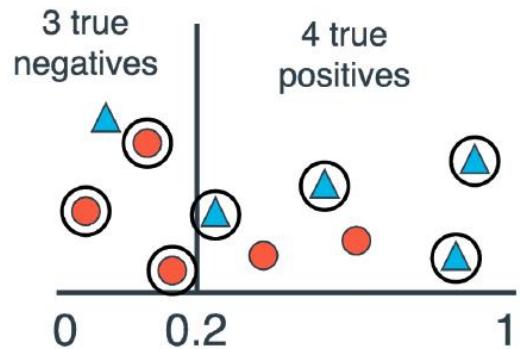
Decision based on ROC

# ROC (Receiver Operating Characteristic)

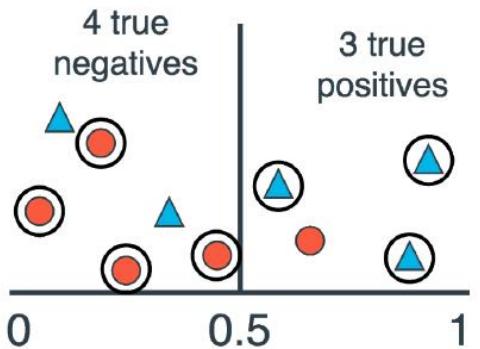


Threshold = 0.5  
Sensitivity = 3/5  
Specificity = 4/5

# ROC (Receiver Operating Characteristic)

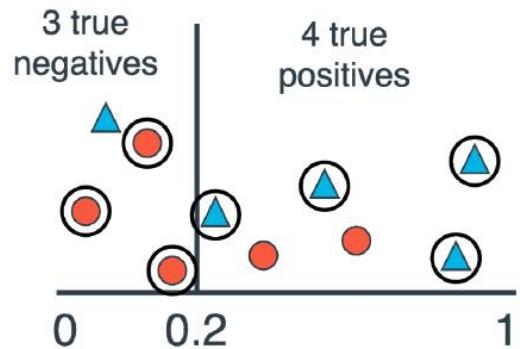


Threshold = 0.2  
Sensitivity = 4/5  
Specificity = 3/5

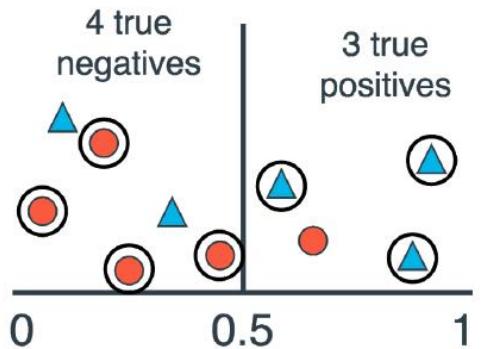


Threshold = 0.5  
Sensitivity = 3/5  
Specificity = 4/5

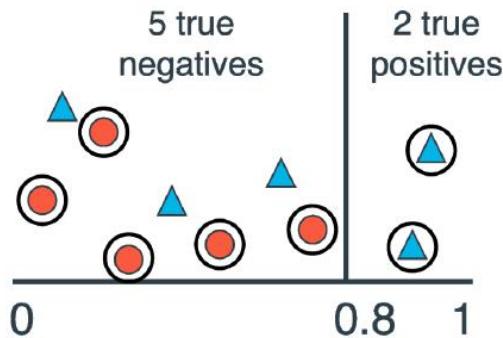
# ROC (Receiver Operating Characteristic)



Threshold = 0.2  
Sensitivity = 4/5  
Specificity = 3/5



Threshold = 0.5  
Sensitivity = 3/5  
Specificity = 4/5



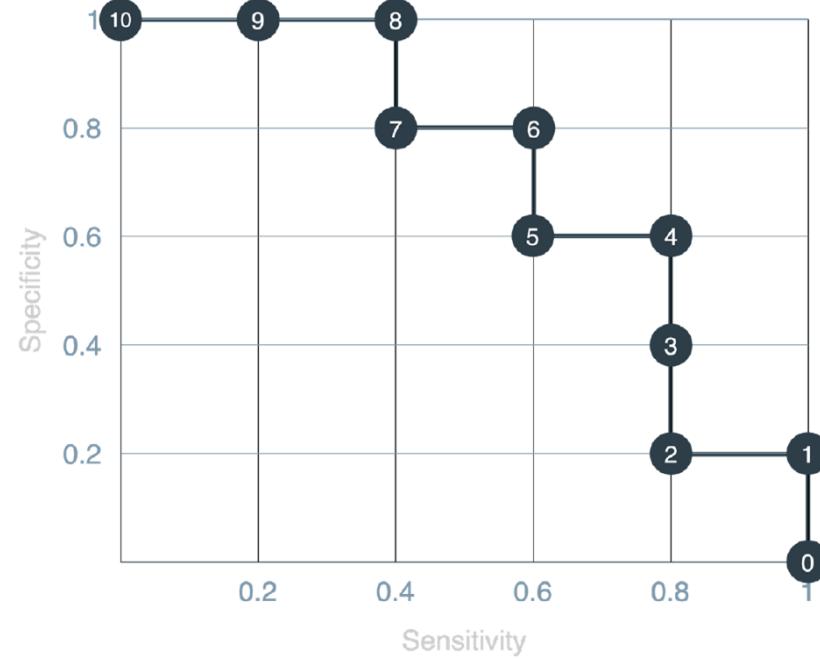
Threshold = 0.8  
Sensitivity = 2/5  
Specificity = 5/5

# ROC (Receiver Operating Characteristic)

Timestep	Threshold	True positives	Sensitivity	True negatives	Specificity
0	0	5	1	0	0
1	0.1	5	1	1	0.2
2	0.2	4	0.8	1	0.2
3	0.3	4	0.8	2	0.4
4	0.4	4	0.8	3	0.6
5	0.5	3	0.6	3	0.6
6	0.6	3	0.6	4	0.8
7	0.7	2	0.4	4	0.8
8	0.8	2	0.4	5	1
9	0.9	1	0.2	5	1
10	1	0	0	5	1

# ROC (Receiver Operating Characteristic)

Timestep	Threshold	True positives	Sensitivity	True negatives	Specificity
0	0	5	1	0	0
1	0.1	5	1	1	0.2
2	0.2	4	0.8	1	0.2
3	0.3	4	0.8	2	0.4
4	0.4	4	0.8	3	0.6
5	0.5	3	0.6	3	0.6
6	0.6	3	0.6	4	0.8
7	0.7	2	0.4	4	0.8
8	0.8	2	0.4	5	1
9	0.9	1	0.2	5	1
10	1	0	0	5	1



# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

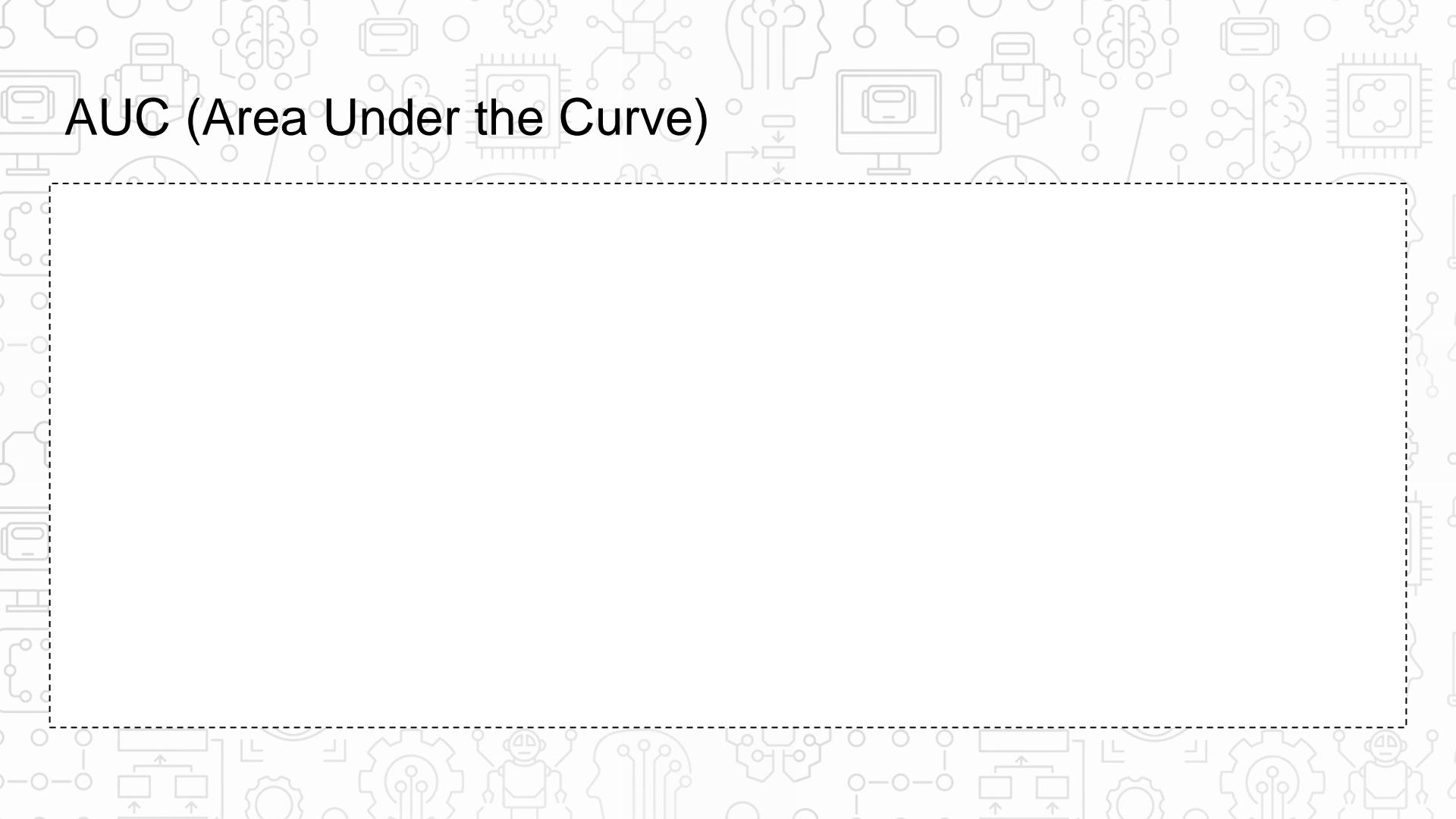
Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

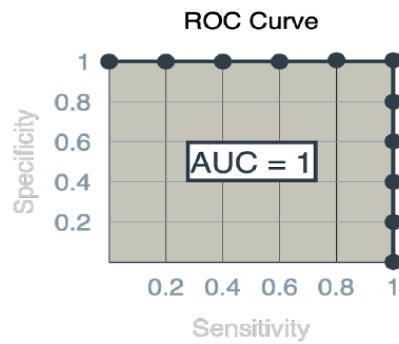
AUC (Area Under the Curve)

Decision based on ROC

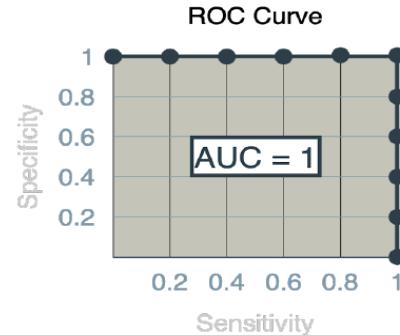
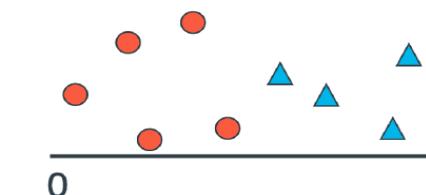
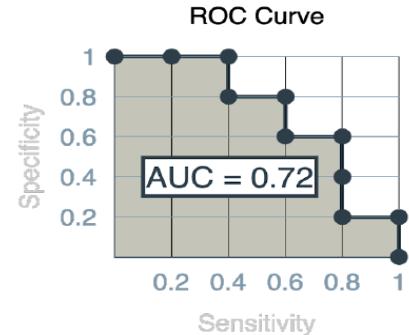
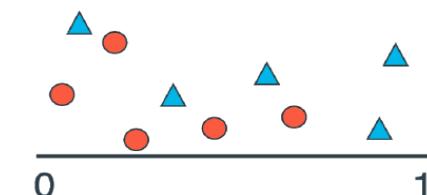
# AUC (Area Under the Curve)



# AUC (Area Under the Curve)



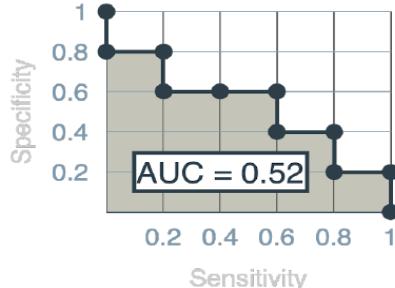
# AUC (Area Under the Curve)



# AUC (Area Under the Curve)

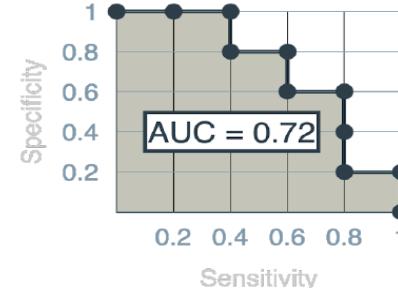
0 1

ROC Curve



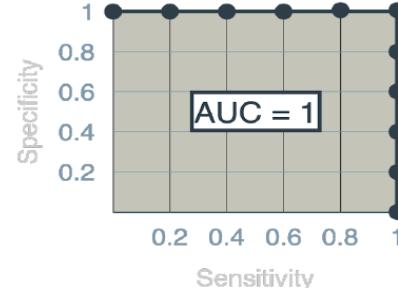
0 1

ROC Curve



0 1

ROC Curve



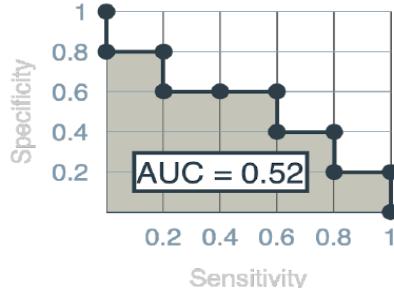
# AUC (Area Under the Curve)

0 1

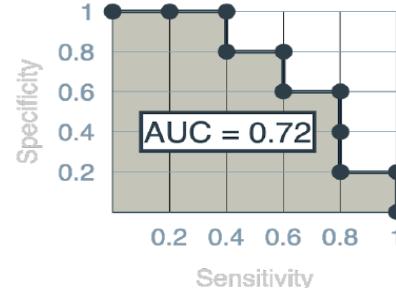
0 1

0 1

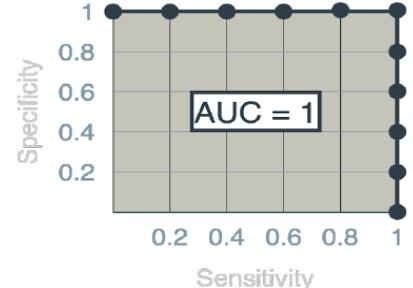
ROC Curve



ROC Curve



ROC Curve



Area Under the Curve tells us how much our model separate the classes.

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

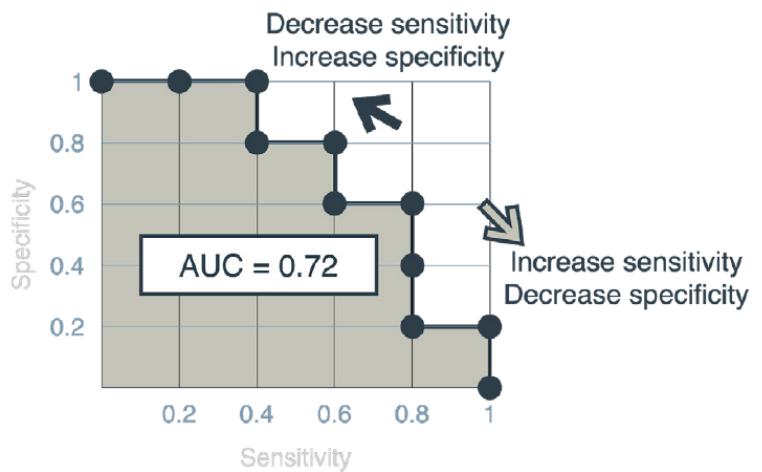
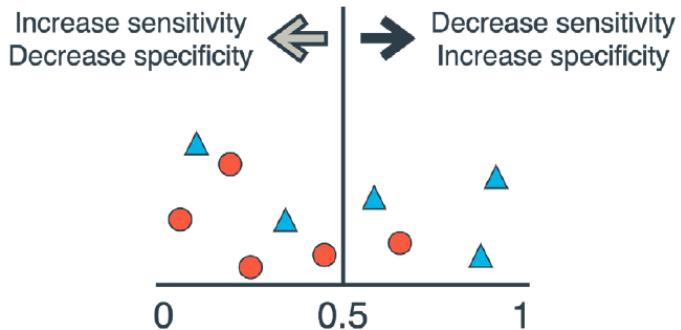
Decision based on ROC

# Decision based on ROC

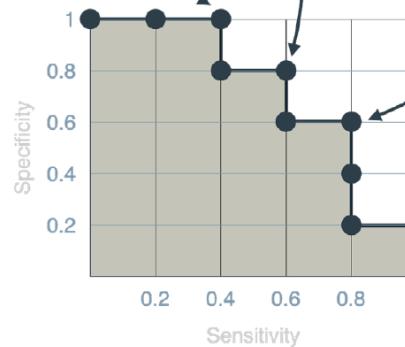
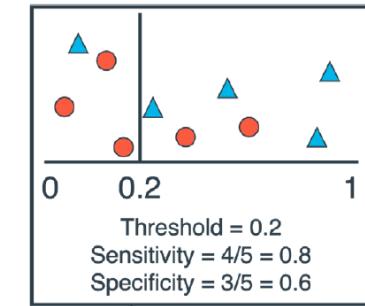
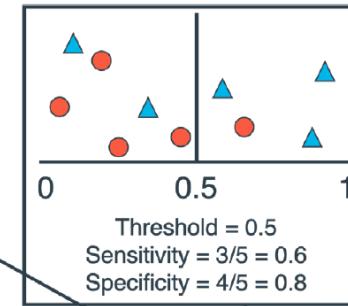
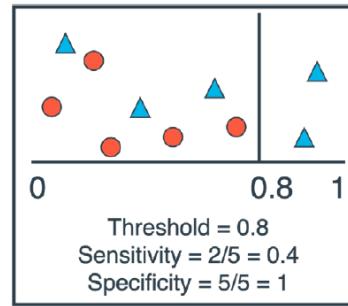
As we increase or decrease the threshold, we change the sensitivity and specificity of the model, and this change is illustrated by moving in the ROC curve.

# Decision based on ROC

As we increase or decrease the threshold, we change the sensitivity and specificity of the model, and this change is illustrated by moving in the ROC curve.



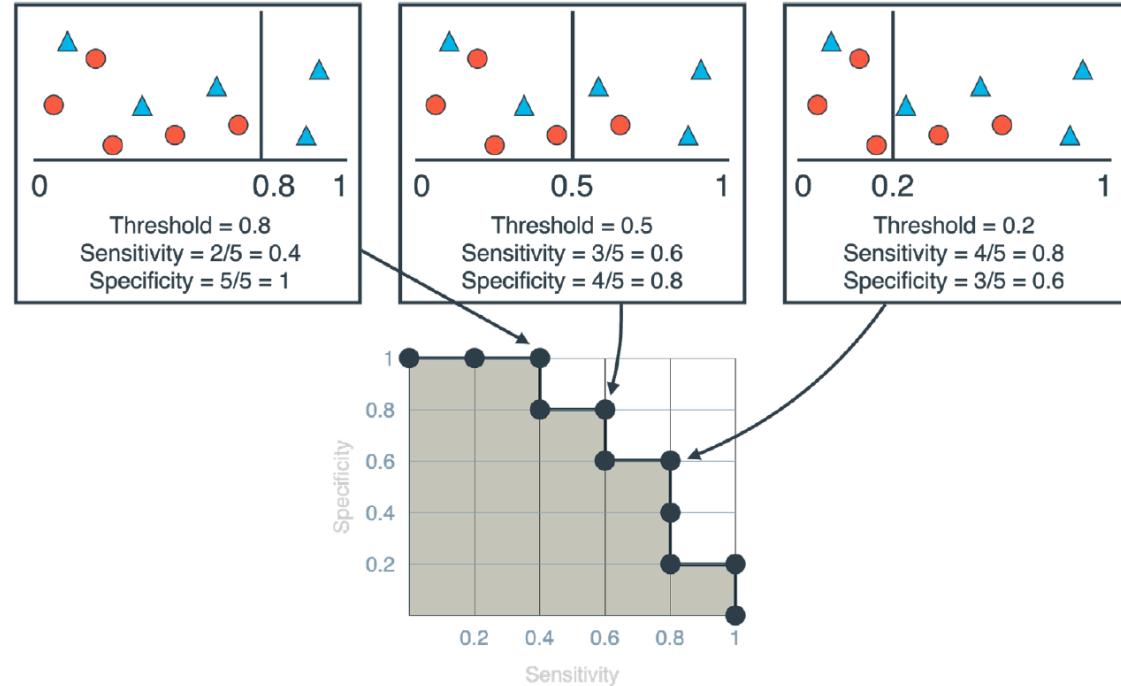
# Decision based on ROC



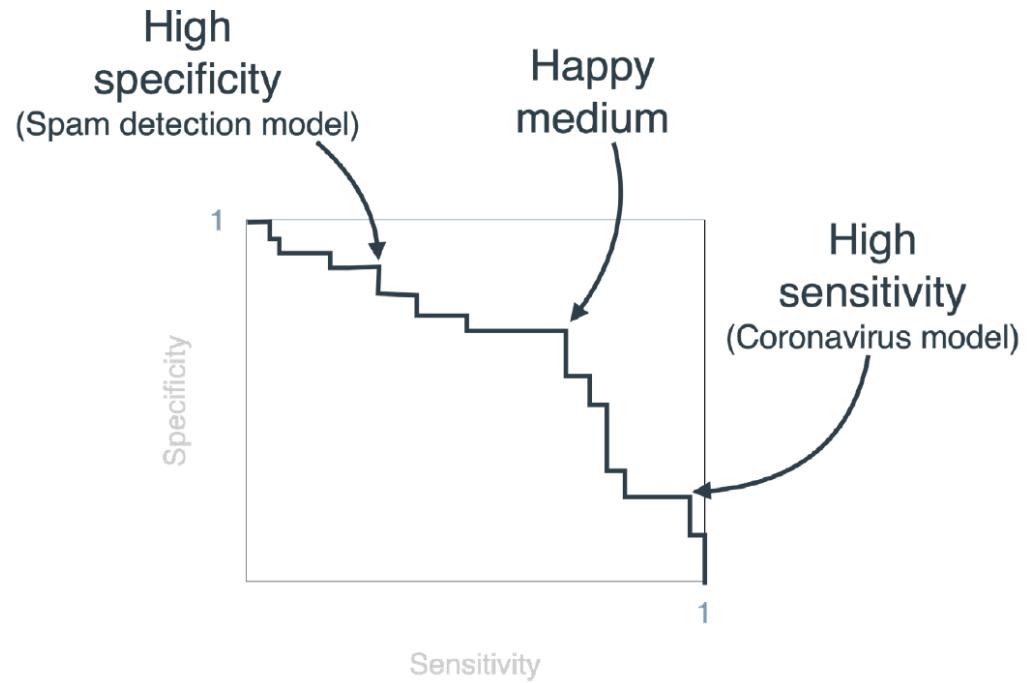
# Decision based on ROC

Mapping the threshold and its ROC.

- The model in the left has a **high threshold, low sensitivity, and high specificity**.
- The model in the middle has **medium values for threshold, sensitivity, and specificity**.
- The model in the right has a **low threshold, high sensitivity, and low specificity**.



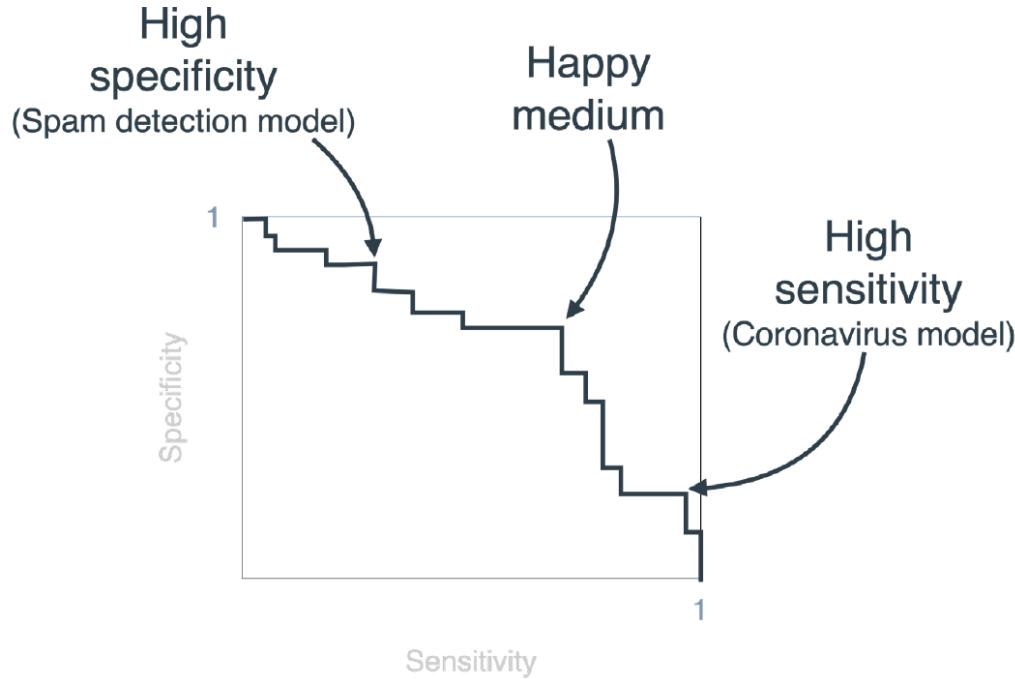
# Decision based on ROC



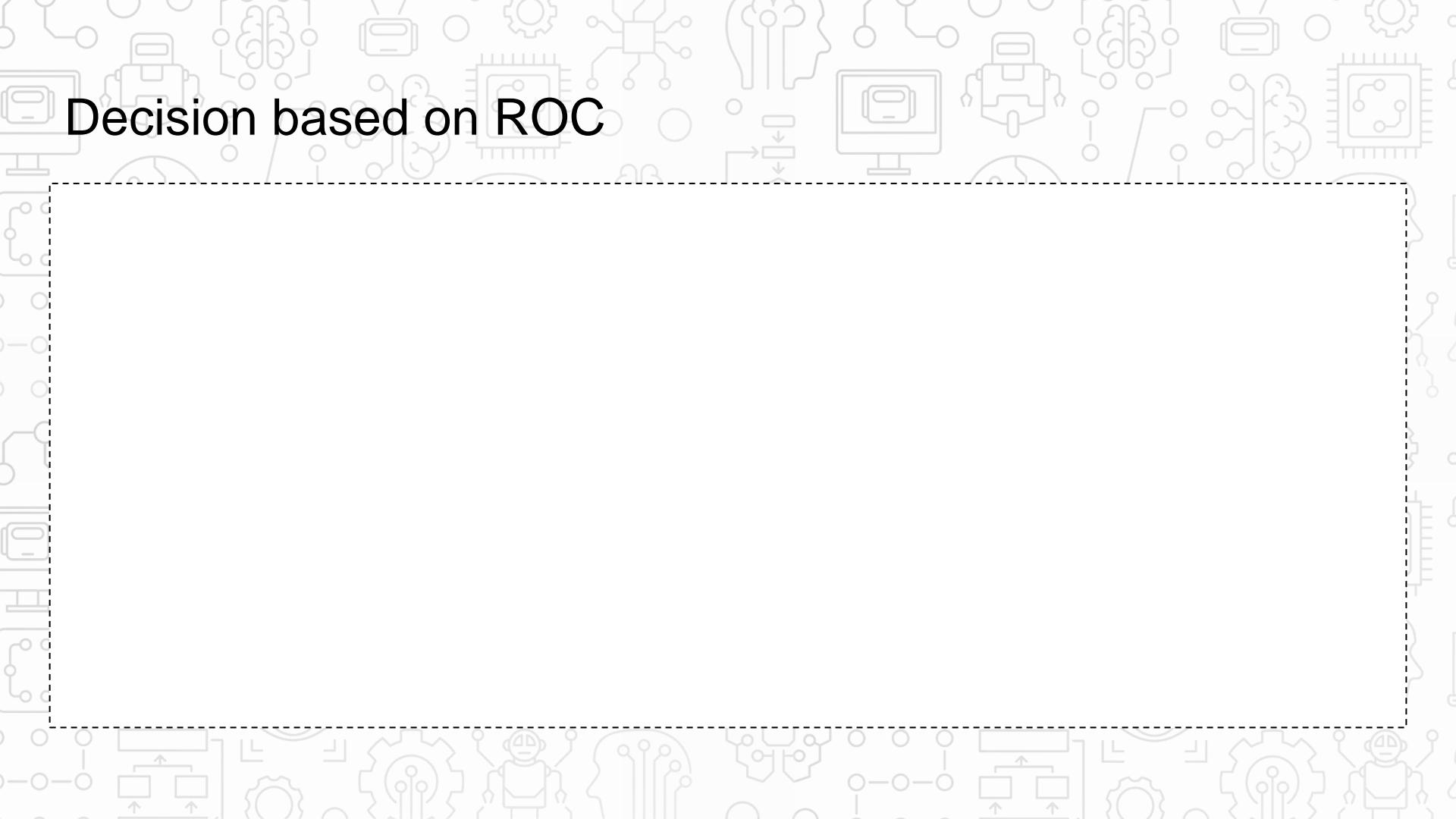
# Decision based on ROC

For problems that needs **high sensitivity** (like coronavirus model), we uses ROC to choose a threshold that achieves that.

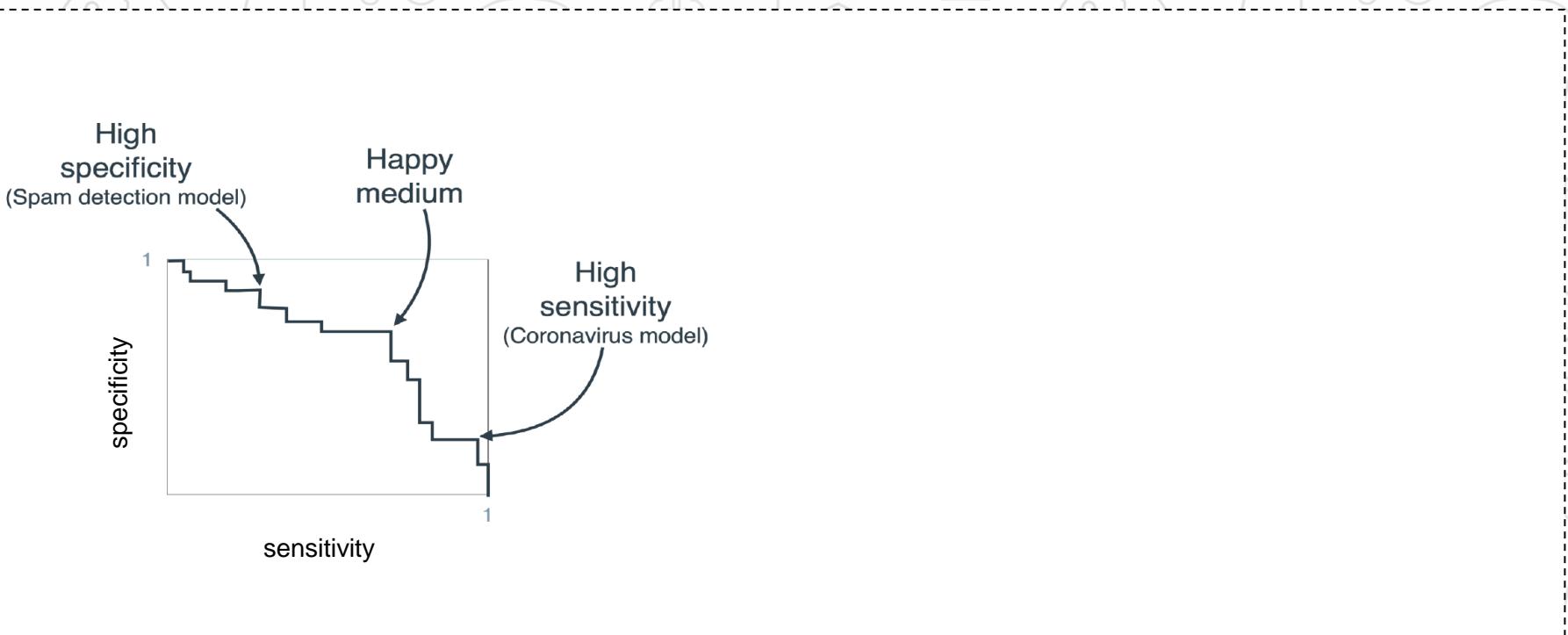
For problems that needs **high specificity** (like spam detector model), we uses ROC to choose a threshold that achieves that.



# Decision based on ROC

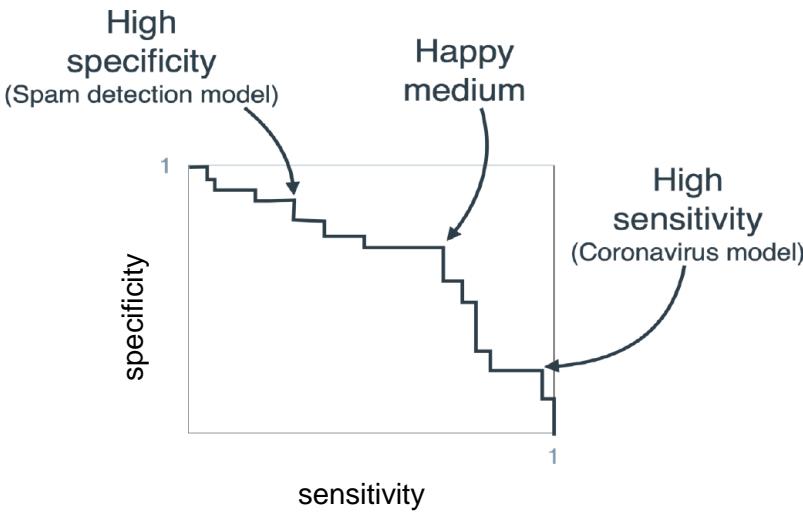


# Decision based on ROC



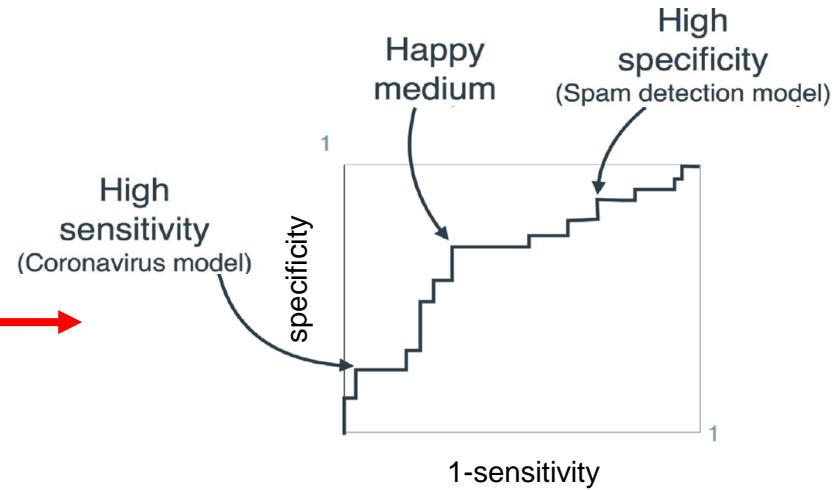
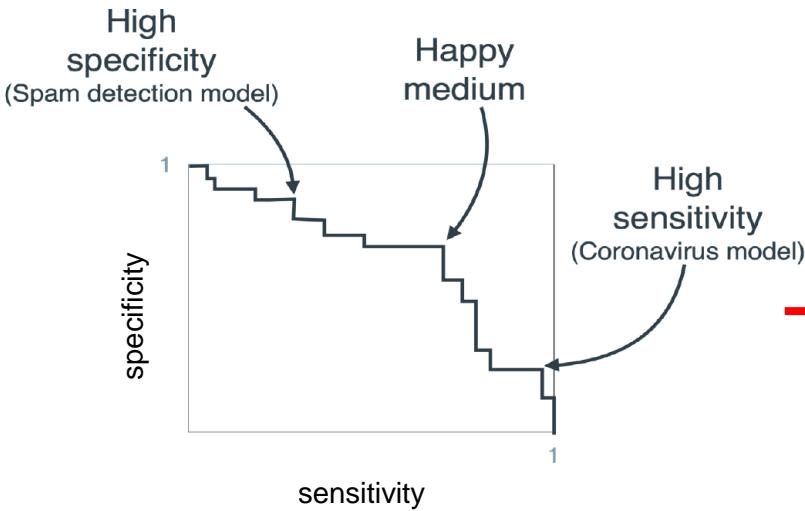
# Decision based on ROC

Usually we plot specificity VS 1-sensitivity

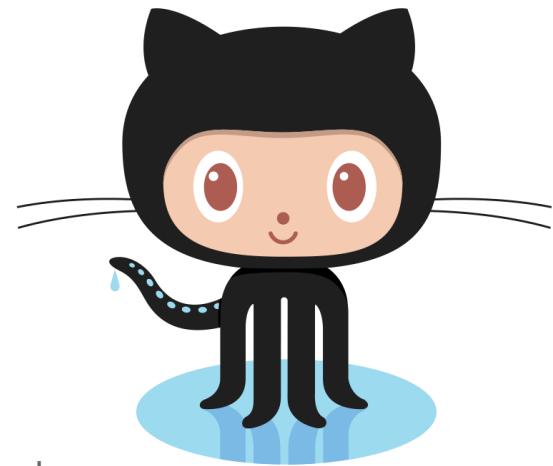


# Decision based on ROC

Usually we plot specificity VS 1-sensitivity



# ROC & AUC



Use colab to open this github notebook:

[“s7s/machine\\_learning\\_1/classification\\_model\\_performance/classification\\_model\\_performance.ipynb”](https://colab.research.google.com/github/s7s/machine_learning_1/blob/main/classification_model_performance/classification_model_performance.ipynb)

# Lecture Overview

Accuracy

False positives & False negatives

Confusion Matrix

Recall & Precision

F-score

Sensitivity and Specificity

ROC (Receiver Operating Characteristic)

AUC (Area Under the Curve)

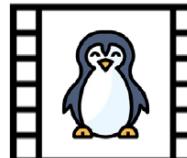
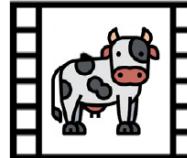
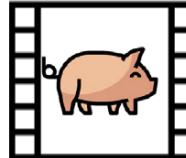
Decision based on ROC

# Exercise 1

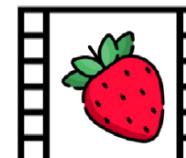
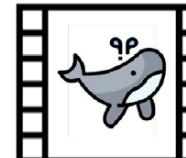
A video site has established that a particular user likes animal videos and absolutely nothing else. You can see the recommendations that this user got when logging into the site.

- a) What is the accuracy of the model?
- b) What is the recall of the model?
- c) What is the precision of the model?
- d) What is the F1-score of the model?
- e) Would you say that this is a good recommendation model?

Recommended



Not recommended



## Exercise2

Find the sensitivity and specificity of the medical model with the following confusion matrix.

	Predicted sick	Predicted healthy
Sick	120	22
Healthy	63	795