ML Concepts
Home | Crash Course
Filter
Crash course
Help Center
ML models
Linear regression (70 min)
Logistic regression (35 min)
Classification (70 min)
Introduction (3 mins)
Thresholds and the confusion matrix (15 min)
Accuracy, recall, precision, and related metrics (15 min)
ROC and AUC (10 min)
Prediction bias (5 min)
Multi-class classification (5 min)
Programming exercise (15 min)
Test your knowledge (10 min)
What's next
Data
Working with numerical data (65 min)
Working with categorical data (30 min)
Datasets, generalization, and overfitting (120 min)

Classification: ROC and AUC

On this page
Receiver operating characteristic curve (ROC)
Area under the curve (AUC)
AUC and ROC for choosing model and threshold
Exercise: Check your understanding

The previous section presented a set of model metrics, all calculated at a single classification threshold value. But if you want to evaluate a model's quality across all possible thresholds, you need different tools.

Receiver-operating characteristic curve (ROC)

The ROC curve is a visual representation of model performance across all thresholds. The long version of the name, receiver operating characteristic, is a holdover from WWII radar detection.

The ROC curve is drawn by calculating the true positive rate (TPR) and false positive rate (FPR) at every possible threshold (in practice, at selected intervals), then graphing TPR over FPR. A perfect model, which at some threshold has a TPR of 1.0 and a FPR of 0.0, can be represented by either a point at (0, 1) if all other thresholds are ignored, or by the following:

Figure 1. ROC and AUC of a hypothetical perfect model.

Area under the curve (AUC)

The **area under the ROC curve (AUC)** represents the probability that the model, if given a randomly chosen positive and negative example, will rate the positive higher than the negative.

The perfect model above, containing a square with sides of length 1, has an area under the curve (AUC) of 1.0. This means there is a 100% probability that the model will correctly rank a randomly chosen positive example higher than a randomly chosen negative example. In other words, looking at the spread of data points below, AUC gives the probability that the model will place a randomly chosen square to the right of a randomly chosen circle, independent of where the threshold is set.

Figure 2. A spread of predictions for a binary classification model. AUC is the chance a randomly chosen square is positioned to the right of a randomly chosen circle.

In more concrete terms, a spam classifier with AUC of 1.0 always assigns a random spam email a higher probability of being spam than a random legitimate email. The actual classification of each email depends on the threshold that you choose.

For a binary classifier, a model that does exactly as well as random guesses or coin flips has a ROC that is a diagonal line from (0,0) to (1,1). The AUC is 0.5, representing a 50% probability of correctly ranking a random positive and negative example.

In the spam classifier example, a spam classifier with AUC of 0.5 assigns a random spam email a higher probability of being spam than a random legitimate email only half the time.

Figure 3. ROC and AUC of completely random guesses.

(Optional, advanced) Precision-recall curve

AUC and ROC work well for comparing models when the dataset is roughly balanced between classes. When the dataset is imbalanced, precision-recall curves (PRCs) and the area under those curves may offer a better comparative visualization of model performance. Precision-recall curves are created by plotting precision on the y-axis and recall on the x-axis across all thresholds.

AUC and ROC for choosing model and threshold

AUC is a useful measure for comparing the performance of two different models, as long as the dataset is roughly balanced. The model with greater area under the curve is generally the better one.

Figure 4. ROC and AUC of two hypothetical models. The curve on the right, with a greater AUC, represents the better of the two models.

The points on a ROC curve closest to (0,1) represent a range of the best-performing thresholds for the given model. As discussed in the Thresholds, Confusion matrix and Precision/recall metrics and trade-offs sections, the threshold you choose depends on which metric is most important to the specific use case. Consider the points A, B, and C in the following diagram, each representing a threshold:

Figure 5. Three labeled points representing thresholds.

If false positives (false alarms) are highly costly, it may make sense to choose a threshold that gives a lower FPR, like the one at point A, even if TPR is reduced. Conversely, if false positives are cheap and false negatives (missed true positives) highly costly, the threshold for point C, which maximizes TPR, may be preferable. If the costs are roughly equivalent, point B may offer the best balance between TPR and FPR.

Here is the ROC curve for the data we have seen before:

Dataset
Separated

Classification threshold

Confusion matrix

	Actually positive	Actually negative
Predicted positive	TP: 90	FP: 60
Predicted negative	FN: 10	TN: 140

Metrics

Accuracy	0.85
Precision	0.85
Recall	0.83

ROC curve

Exercise: Check your understanding

In practice, ROC curves are much less regular than the illustrations given above. Which of the following models, represented by their ROC curve and AUC, has the best performance?

This model has the highest AUC, which corresponds with the best performance.

Correct answer.

Which of the following models performs worse than chance?

This model has an AUC lower than 0.5, which means it performs worse than chance.

Correct answer.

(Optional, advanced) Bonus question

Which of the following changes can be made to the worse-than-chance model in the previous question to cause it to perform better than chance?

Reverse the predictions, so predictions of 1 become 0, and predictions of 0 become 1.

For a binary classifier reliably pick examples in the wrong classes more often than chance, switching the class label immediately makes its predictions better than chance without having to retrain the model.

Correct answer.

Have it always predict the positive class.

Have it always predict the negative class.

Imagine a situation where it's better to allow some spam to reach the inbox than to send a business-critical email to the spam folder. You've trained a spam classifier for this situation where the positive class is spam and the negative class is not spam. Which of the following points on the ROC curve for your classifier is preferable?

Point A

In this use case, it's better to minimize false positives, even if true positives also decrease.

Correct answer.

Point B

Point C

Key terms:

- Area under the ROC curve (AUC)
- ROC curve

Help Center

Was this helpful?
Send feedback

Except as otherwise noted, the content of this page is licensed under the Creative Commons Attribution 4.0 License, and code samples are licensed under the Apache 2.0 License. For details, see the Google Developers Site Policies. Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2025-06-02 UTC.