

Watch video on [ROC Curves and Area Under the Curve](#) (14 minutes), and be prepared to **discuss it in class** on Wednesday. (Feel free to play with the [visualization](#) shown in the video. Here are some questions to think about:

- What is the difference between the [predict](#) and [predict_proba](#) methods in scikit-learn?
- If you have a classification model that outputs predicted probabilities, how could you convert those probabilities to class predictions?
- Why are predicted probabilities (rather than just class predictions) required to generate an ROC curve?
- Could you use an ROC curve for a regression problem? Why or why not?
- What's another term for True Positive Rate?
- If I wanted to increase specificity, how would I change the classification threshold?
- Is it possible to adjust your classification threshold such that both sensitivity and specificity increase simultaneously? Why or why not?
- What are the primary benefits of ROC curves over classification accuracy?
- What should you do if your AUC is 0.2?
- What would the plot of reds and blues look like for a dataset in which each observation was a credit card transaction, and the response variable was whether or not the transaction was fraudulent? (0 = not fraudulent, 1 = fraudulent)
- What's a real-world scenario in which you would prefer high specificity (rather than high sensitivity) for your classifier?