Model Evaluation

Confusion Matrix, Accuracy, Recall, Precision, f1-score, ROC, AUC

Confusion matrix

- Actual 1, Predicted 1 ==> True Positive (TP)
- Actual 1, Predicted 0 ==> False Negative (FN)
- Actual 0, Predicted 0 ==> True Negative (TN)
- Actual 0, Predicted 1 ==> False Positive (FP)

- Accuracy = Correct Prediction / Total observation
- = (TP + TN) / (TP + TN + FP + FN)

Confusion Matrix

- To evaluate the performance of your model, you collect
  - **10,000** manually classified transactions out of which
    - **300** are *fraudulent* transactions and
    - **9,700** *non-fraudulent* transactions.
- Run your classifier on every transaction and predict the class label
  - *fraudulent* or non-*fraudulent*
  - and summarize the results in a *confusion matrix*

Trick:
TP: Truly predicted Positive
TN: Truly predicted Negative
FP: Falsely predicted Positive
FN: Falsely predicted Negative

Confusion Matrix

- **True Positive (TP=100)** – the model *correctly* predicts the *positive (fraudulent)* class.
- **True Negative (TN=9,000)** – the model *correctly* predicts the *negative (non-fraudulent)* class.
- **False Positive (FP=700)** – the model *incorrectly* predicts the *positive (fraudulent)* class
- **False Negative (FN=200)** – the model *incorrectly* predicts the *negative (non-fraudulent)* class.

	Predicted Negative	**Predicted Positive**	
Actual Negative	True Negative 9,000	False Positive 700	Actual non-Fraudulent (Negative)
Actual Positive	False Negative 200	True Positive 100	Actual Fraudulent (Positive)
	Predicted as non-Fraudulent (Negative)	Predicted as Fraudulent (Positive)	

Accuracy

		Predicted N	Predicted P
Actual	N	TN 9,000	FP 700
	P	FN 200	TP 100

- **Accuracy**: Correctness of predictions

$$Accuracy = \frac{True}{True+False} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{100+9{,}000}{100+9{,}000+700+200} = \frac{9{,}100}{10{,}000} = 0.91$$

		Predicted N	Predicted P
Actual	N	TN 9,700	FP 0
	P	FN 300	TP 0

- In case we predict **all transactions as non-fraudulent:**

$$Accuracy = \frac{True}{True+False} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+9{,}700}{100+9{,}000+700+200} = \frac{9{,}700}{10{,}000} = 0.97$$

Recall, Precision, f1 score

Actual \ Predicted	N	P
N	TN 9,000	FP 700
P	FN 200	TP 100

- **Recall:** What proportion of positives are being predicted correctly?
  - The classifier caught 33.3% of the fraudulent transactions (True +ve / Total +ve)

$$Recall(TruePositiveRate) = \frac{TP}{TP+FN} = \frac{100}{100+200} \approx 0.333$$

- **Precision:** What proportion of predicted positives are correct?
  - When your classifier predicts that a transaction is *fraudulent*, **only 12.5% of the time your classifier is correct** (***True +ve / Total predicted +ve***)

$$Precision = \frac{TP}{TP+FP} = \frac{100}{100+700} = 0.125$$

- **f1 Score** combines Recall and Precision to one performance metric
  - Harmonic mean of Recall and Precision
  - This score takes both false positives and false negatives into account

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} = 2 * \frac{0.333 * 0.125}{0.333 + 0.125} \approx 0.182$$

Recall, Precision, f1 score

		Predicted N	Predicted P
Actual	N	TN 9,700	FP 0
Actual	P	FN 300	TP 0

- **Recall:** What proportion of positives are being predicted correctly?
  - TP / (TP + FN)
  - Which calculates to 0 !!

- **Precision:** What proportion of predicted positives are correct?
  - TP / (TP + FP)
  - Which again calculates to 0 !!

- **f1 Score** combines Recall and Precision to one performance metric
  - Harmonic mean of Precision and Recall.
  - This score takes both false positives and false negatives into account
  - 2 * (Recall * Precision) / (Recall + Precision)

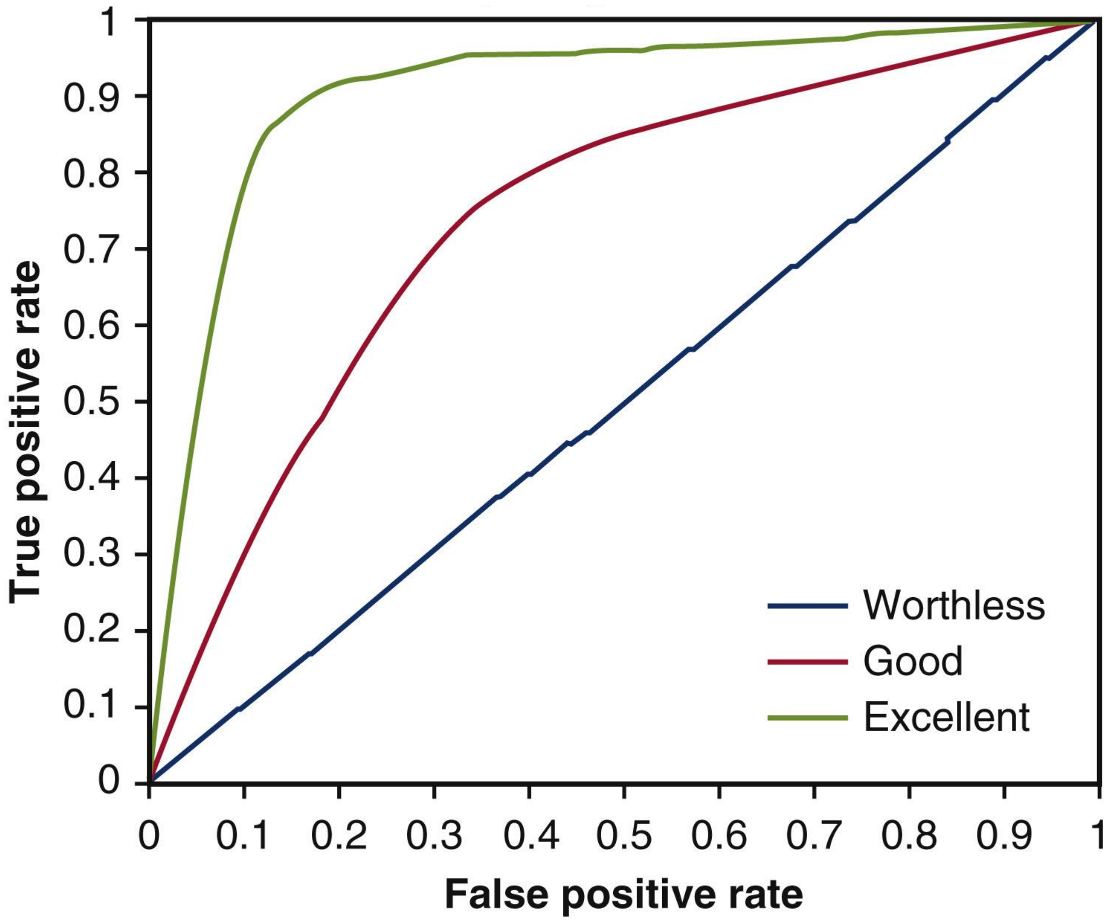
Intuition

- **Accuracy** - % of correct prediction = TP + TN / TP+FN+TN+FP
- **Recall** - % of positives predicted correctly = TP / TP + FN
- **Precision** - % of predicted positives are correct = TP / TP + FP
- **F1** – Harmonic mean of Precision and Recall

  = 2* (Recall*Precision) / (Recall + Precision)

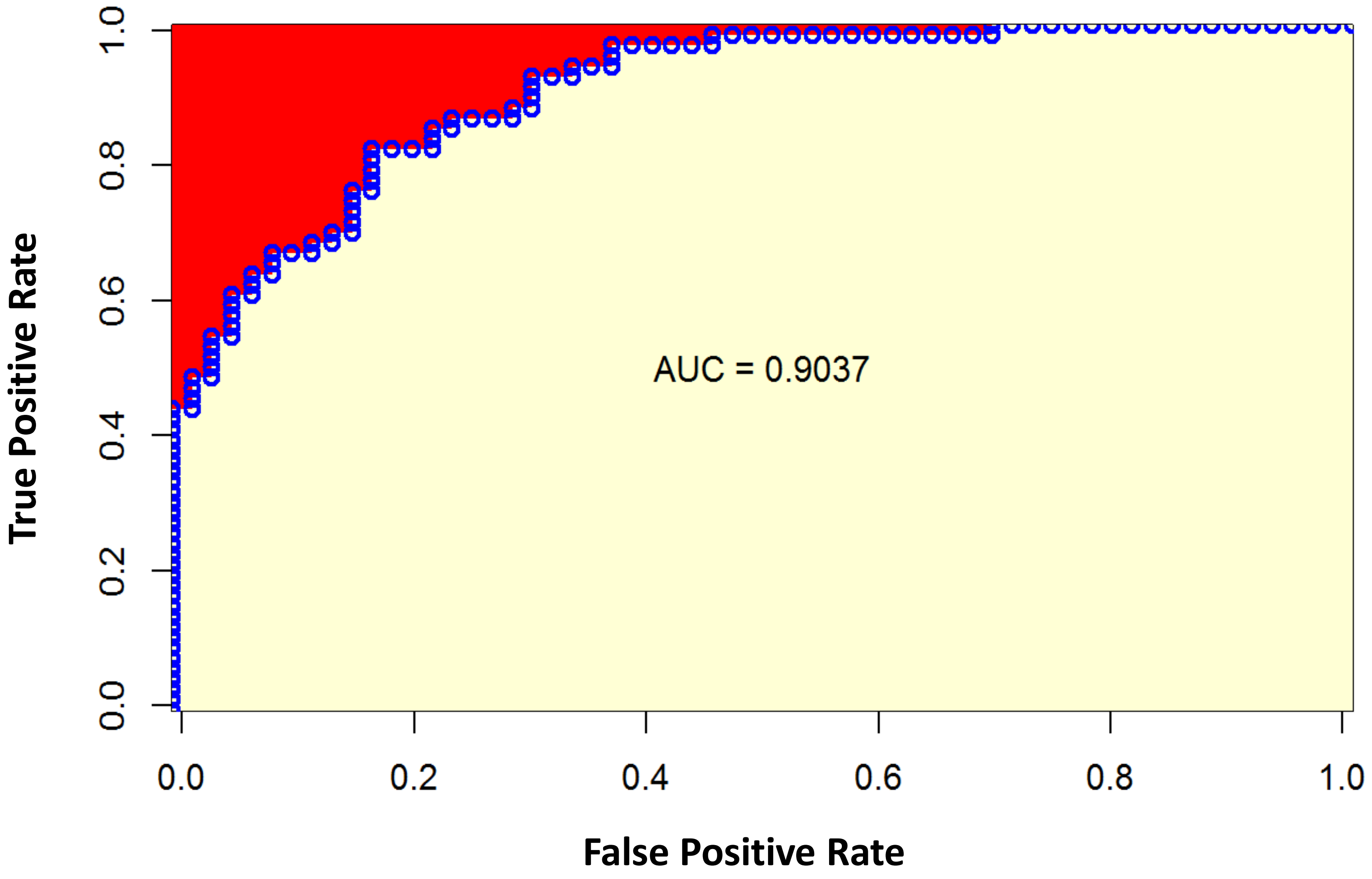
ROC (Receiver Operating Characteristics)

- **True Positive Rate (TPR)** = % of positives predicted correctly = Recall = TP / TP + FN
- **False Positive Rate (FPR)** - % error in predicting negatives = FP / FP+TN



- ROC Curves are used to see how well your classifier can separate positive and negative examples
- To be able to use the ROC curve, your classifier should be able to rank examples such that the ones with higher rank are more likely to be positive (*fraudulent*).
- As an example, Logistic Regression outputs *probabilities*, which is a score that you can use for ranking

AUC (Area Under the ROC)



- The model performance is determined by looking at the area under the ROC curve (or AUC).
- An excellent model has AUC near to the 1.0, which means it has a good measure of separability