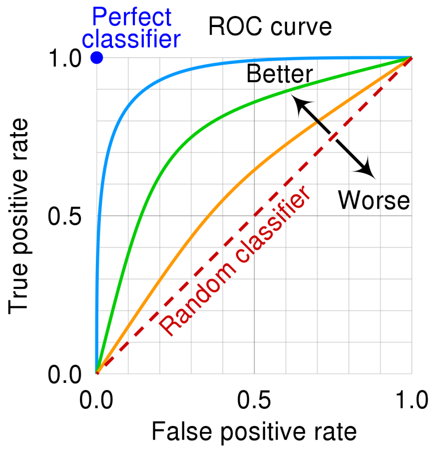
Evaluation Matrices

- **Classification**
  - Confusion Matrix
    - determine the performance of classifier, containing the info about actual and predicted classifications

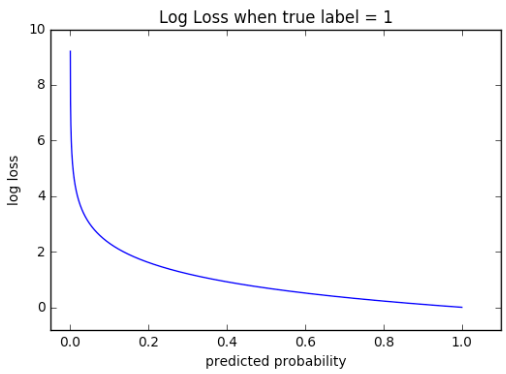
		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	Negative Predictive Value $\frac{TN}{(TN+FN)}$	Accuracy $\frac{TP+TN}{(TP+TN+FP+FN)}$

      - $Sensitivity/Recall = \frac{TP}{TP+FN}$, positive examples labeled as positive by classifier.
        - *When it's actually yes, how often does it predict yes?*
      - $Specificity = \frac{TN}{FP+TN}$, negative examples labeled as negative by classifier.
        - *When it's no, how often does it predict no?*
      - $Precision = \frac{TP}{TP+FP}$, shows correctness achieved in positive prediction.
        - *When it predicts yes, how often is it correct?*
      - $Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$, the proportion of the total number of predictions that are correct.
        - *Overall, how often is the classifier correct?*
      - $F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall}$
        - a weighted average of the recall (sensitivity) and precision.
  - Receiver Operating Characteristics(ROC) and AUC
    - **y-axis:** true positive rate(sensitivity), how many correct positive results occur among all positive samples available during the test; **x-axis**: false positive rate(1-specificity), defines how many incorrect positive results occur among all *negative* samples available during the test.



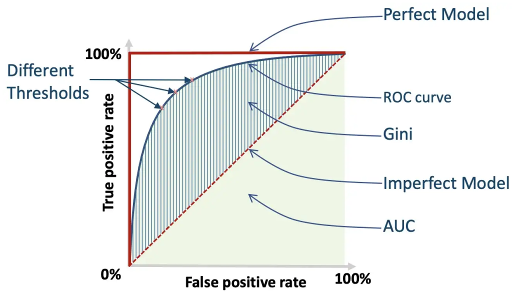
- ROC: probability curve.
- AUC:It tells how much the model is capable of distinguishing between classes. The higher, the better. [0,1]

- Log Loss
  - definition: negative average of the log of corrected predicted probabilities for each instance. The more the predicted probability diverges from the actual value, the higher is the log loss value.
  - Function: $-\frac{1}{N} * \sum_{i}^{N} y_i * log(p(y_i)) + (1 - y_i) * log(1 - p(y_i))$
    - $Logloss = \frac{1}{N} \sum_{i}^{N} logloss_i$
    - N : number of observations
  - Graph:



    - the lower, the better.
    - Sometimes, we set a baseline/threshold value of logloss, with a constant probability =xx% of data with class 1 observations.
    - Higher imbalance in a dataset, lower baseline log-loss score of the dataset

- Gini Coefficient
  - Function: $Gini = 2 * AUC - 1$
  - Gini>60%, the model is good enough. The range is [-1,1]
  - Definiton: the ratio between area between the ROC curve and the diagnol line & the area of the above triangle.



- **Regression**
  - Root Mean Squared Error(RMSE)
    - $RMSE = \sqrt[2]{\frac{1}{n}\sum_i^n(y_i - \hat{y})^2}$
      - y_i:actual value
      - $\hat{y}$:predicted value
      - n: # of observations
  - R-Squared & Adjusted R-Squared
    - $R^2 = 1 - \frac{\sum_i(y_i - \hat{y}_i)^2}{\sum_i(y_i - \bar{y})^2}$
      - the higher, the better
    - $Adjusted\ R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$
      - N: total sample size
      - p : number of independent variable(feature)
  - Mean Absolute Error(MAE)
    - $MAE = \frac{1}{n}\sum_i^n|y_i - \hat{y}|$
  - Mean Squared Error(MSE)
    - $MSE = \frac{1}{N}\sum_j^N(y_j - \hat{y}_j)^2$
    - measures the average of the squares of the errors(differences between estimated value and actual value)
- Reference:
  - https://en.wikipedia.org/wiki/Receiver_operating_characteristic