Model Comparison

Classification criteria - based on fixed cut-off

- Confusion matrix made by Test data

	predict 0	predict 1
Actual 0	True Negative	False Positive
Actual 1	False Negetive	True Positive

- **Accuracy** : $\frac{TN+TP}{TN+TP+FN+FP}$

  Accuracy can measure overall classification power for classification model.

  Which can measure the general performance of model intuitively

- **Sensitivity** (= Recall) : $\frac{TP}{TP+FN}$

  Sensitivity measures 'How well Actual True is detected by model'

  $P(\text{prediction is correct}|A=1)$

- **Specificity** : $\frac{TN}{TN+FP}$

  Specificity measures 'How well Actual False is detected by model'

  $P(\text{prediction is correct}|A=0)$

- **Precision** : $\frac{TP}{TP+FN}$

  Precision measures 'How precisely model predict actual 1'

  $P(\text{prediction is correct}|P=1)$

- **F1 - score** : $\frac{2 \cdot Sensitivity \cdot precision}{Sensitivity + Precision}$

  F1 score is Harmonic mean with equal weight between Sensitivity and Precision

  In many cases, the objective of model is to classify True value (e.g - defective product, Cancer patient, etc.)

  F1 score measure 'How well and How precisely model predict actual True value'

  F1 score for comparison must be made by test data. So the Higher F1 score the model has, The better performance the model made. (which means there is no concern about overfitting)

When Data is unbalanced

- If Data consists of 99% False and 1% True, then a model which predict every value as False has 0.99 Accuracy

  $\rightarrow$ this is not good!

- Because of this, We have to consider Sensitivity and precision when data is unbalanced.

Classification criteria - based on ROC

- When cut-off is changed, performance of model is also changed.
- Comparison criteror based on fixed cut-off can not compare models in entire cut-offs

  $\rightarrow$ We need other criteria
- Sensitivity and Specificity has inverse relationship when cutoff change.

  cutoff increases $\rightarrow$ TN increase $\rightarrow$ Specificity increase

  cutoff increases $\rightarrow$ FN increase $\rightarrow$ Sensitivity decrease

  cutoff decreases $\rightarrow$ TP increase $\rightarrow$ Sensitivity increase

  cutoff decreases $\rightarrow$ FP increase $\rightarrow$ Specificity decrease
- By this property, We can make gragh drawn by Sensitivity and Specificity on each cut-off value.
- When on same specificity, The model which have higher Sensitivity is considered as the better model.
- Area Under ROC curve is quantitative criteria with this mechanism.
- Model not affected by cut-off value has 0.5 AUROC value.

Regression criteria

- Predictive $R^2 : 1 - \frac{\sum(y_i - \hat{y}_i)}{\sum(y_i - \bar{y})}$

  Which is same as $corr(y, \hat{y})^2$

  Predictive R^2 can be distorted when error has tendency
- Mean Absolute Error (MAE) : $\frac{1}{n} \sum |y_i - \hat{y}_i|$

  Order of MAE and Predictive R^2 are usually same.
- Mean Absolute Percentage Error(MAPE) : $\frac{1}{n} \sum \frac{|y_i - \hat{y}_i|}{y_i}$

  MAPE considers the scale of y
- Mean squared Error (MSE): $\frac{1}{n} \sum (y_i - \hat{y}_i)^2$
- Rooted MSE (RMSE) : $\sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$