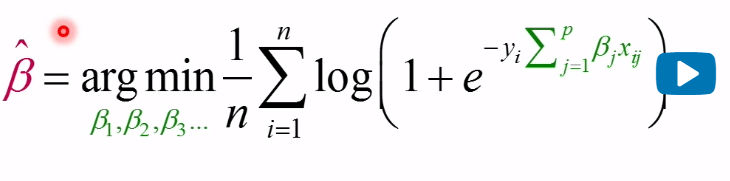
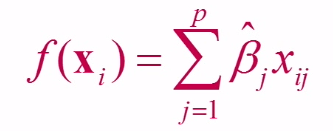
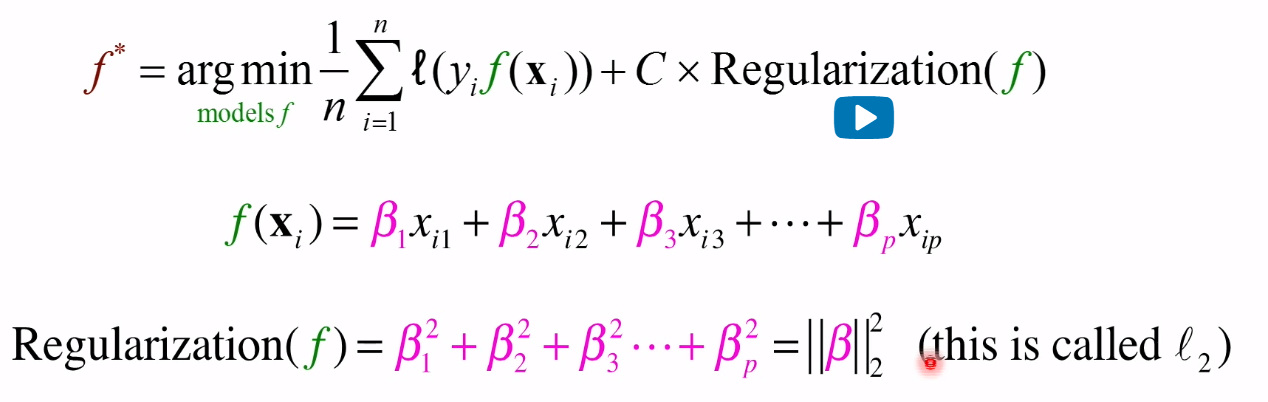
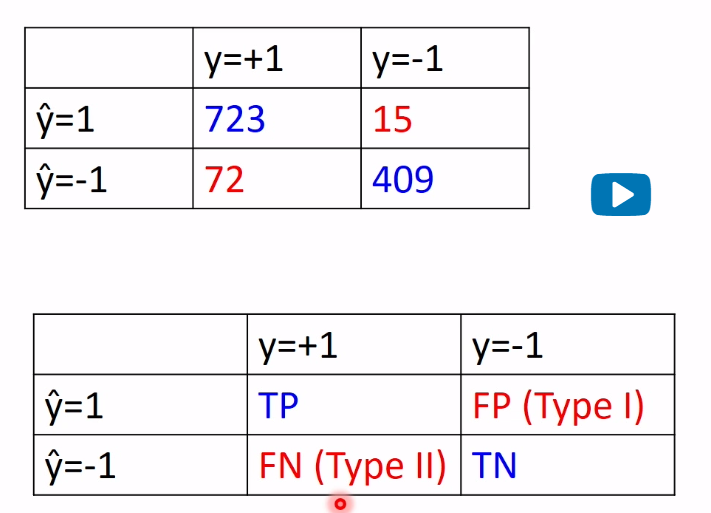
# LOGISTIC REGRESSION

* Step 1: Split data randomly into training and test sets
* Step 2: Estimate coefficients / Train model
* 
* Step 3: Score model: Compute score for each xi in the test set
* 
* Step 4: Evaluate model
* Improving performance through regularization
* 

# CLASSIFICATION

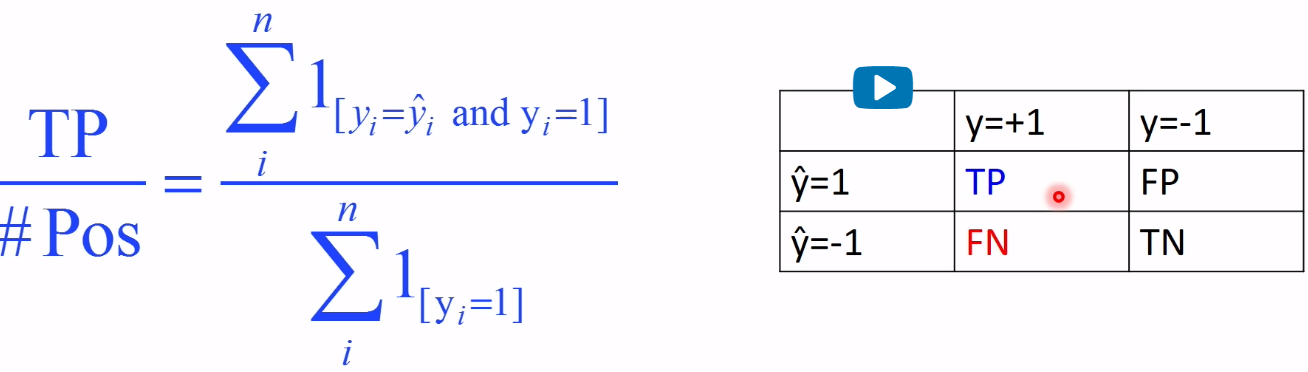
* True Negative – Cuando la etiqueta es negativa y lo predecimos como negativo
* True Positive – Cuando la etiqueta es positiva y lo predecimos positivo
* False Positive – Cuando la etiqueta es negativa y predecimos positiva (Error Type 1)
* Falso Negativo – Cuando la etiqueta es positiva y predecimos negativo (Error Type 2)

## CONFUSION MATRIX

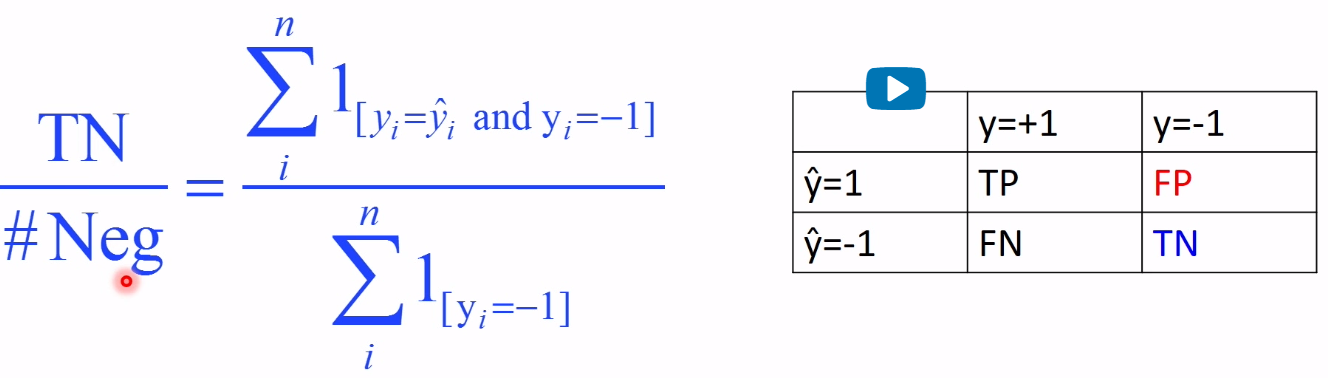


## Evaluation Measures for Classifiers

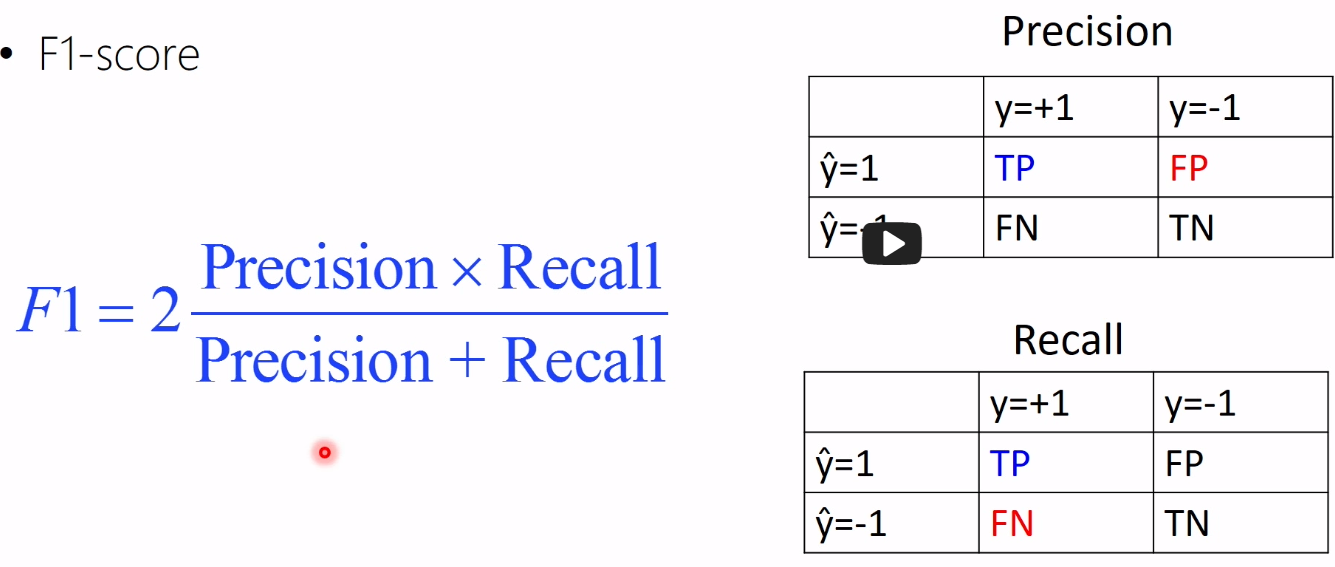
* True Positive Rate (TPR), Sensitivity, Recall.



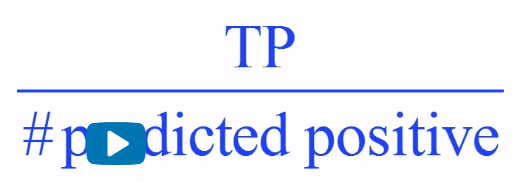
* True Negative Rate (TNR), Sensitivity.



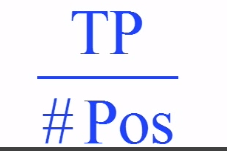
* F1-score



* Precision@N for a search query: Of the top N pages received, how many relevant?



* Recall@N for a search query: If there are N: #Pos relevant webpages, what fraction did our query return?

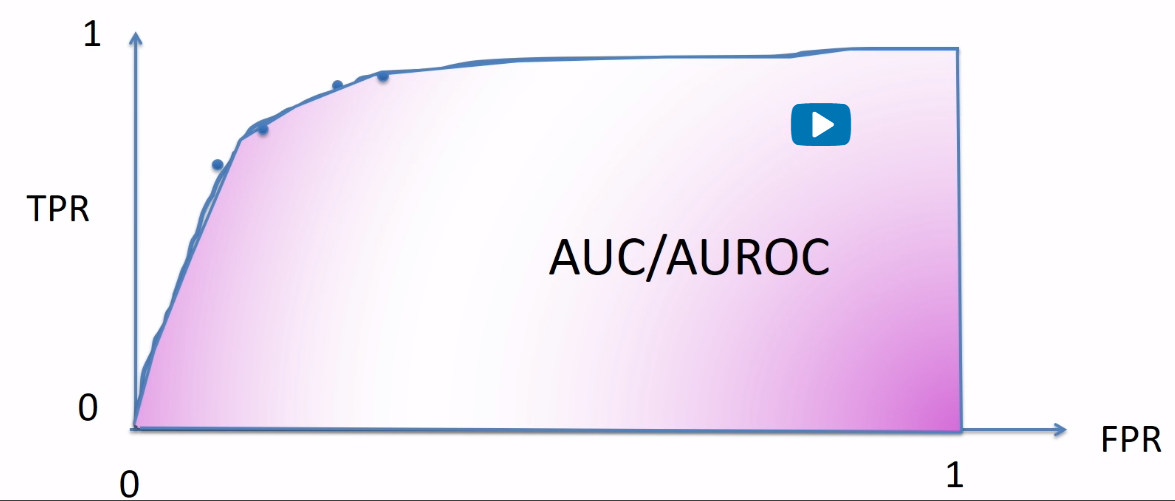


## Which one?

* Standard ML: Missclassification Error (or Accuracy)
* Doctors: Sensitivity and Specificity (TPR and TNR)
* Info Retrieval: Precision & Recall F1-score

# ROC Curves:

* For a particular False Positive Rate (FPR), what is the True Positive Rate (TPR)?



## Evaluation:

* Many ways to evaluate a model:
  + Confusion matrix (TP, TN, FP, FN)
  + Accuracy / misclassification error
  + Precision, Recall, F1-score
  + ROC curves, AUC/AUROC

# ROC CURVE ALGORITHMS

* ROC Curves can be produced in 2 ways:
  + Using a single real-valued classifier. In that case the ROC curve evaluates the **classifier.**
  + Using a single algorithm and sweeping the imbalance parameter across the full range. In that case, the ROC curve evaluates the **algorithm.**

