Feb'03' 2023

Thursday, February 2, 2023 10:08 PM

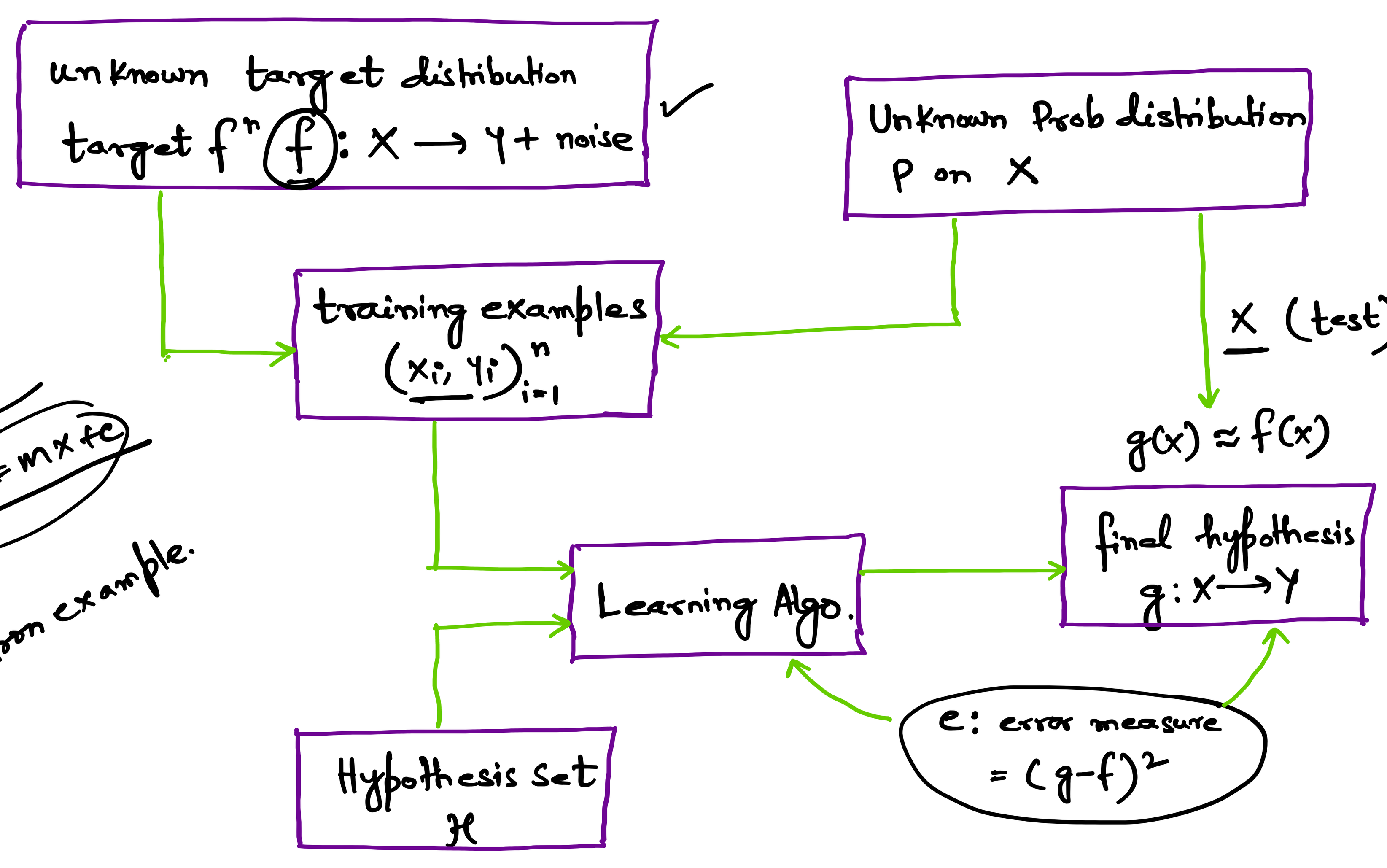
① Cross-Validation

② Evaluation metrics

③ Softmax Vs Logistic Regression.

$f: X \rightarrow Y$ $\qquad y + \underline{\text{noise}}$

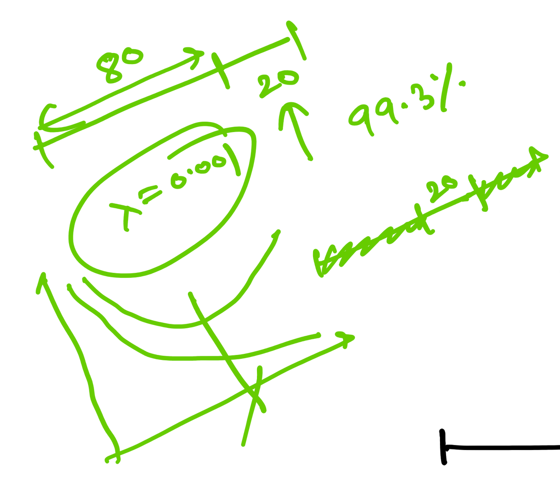
① Cross-Validation:

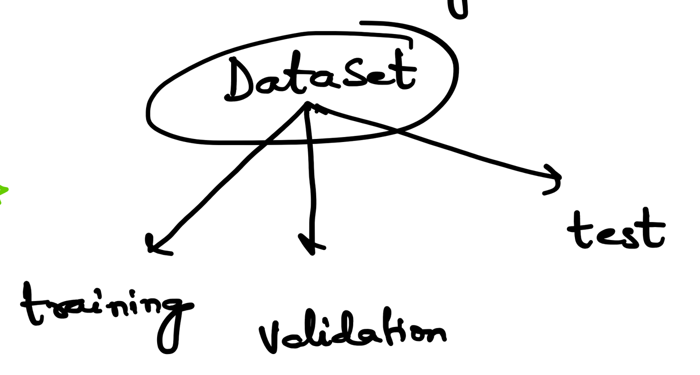


$y = mx + c$

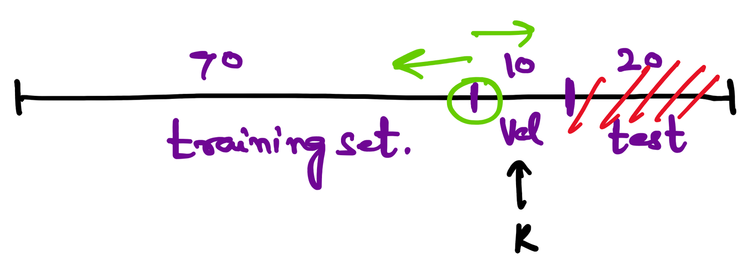
perceptron example.

⇒ We get a model: How good it is.





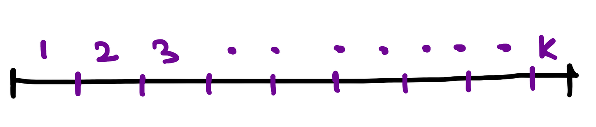
Exam Analogy:
- Lecture, HW,
- Finals.



you need K to be both large & small.

⇒ cross-Validation comes to rescue

tune hyper-parameter.

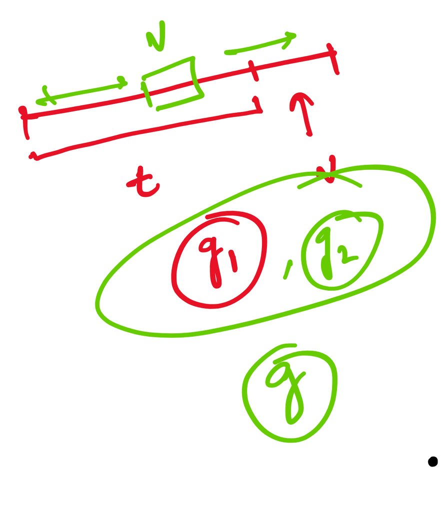


Combine training + Validation.

split data into K fold.

use: (K−1) fold for training.

1 " " Validation.



Finalize hyper-parameter.

Entire training + Validation set to repeat the model. } g

∴ Entire set for training & Validation.

② Evaluation Metric:

JP Margan Example: $\boxed{X} \rightarrow \boxed{M} \longrightarrow 99\%$.

		Actual	
		+ve	−ve
Predicted	+ve	True +ve TP	False +ve FP
	−ve	False −ve FN	True −ve TN

$P = TP/(TP+FP)$

$R = \dfrac{TP}{(TP+FN)}$

(i) **Accuracy:** What % of the time model is classifying correctly.

(Well balanced No skew or class imbalance).

$$\therefore \ Acc = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

(ii) **Precision:** what proportion of predicted positives are truly positives.?

(When we want to be very sure of our prediction.)

$$P = \frac{TP}{(TP+FP)}$$

(iii) **Recall:** What proportion of actual positives are correctly classified.

(when want to capture as many +ve as possible)

$$R = \frac{TP}{(TP+FN)}$$

(iv) **F1 score:** tradeoff b/w Precision & Recall. It is harmonic mean of Precision & Recall.

$$\frac{1}{F_1} = \frac{1}{2}\left(\frac{1}{P} + \frac{1}{R}\right)$$

$$\Rightarrow F_1 = 2\left(\frac{PR}{P+R}\right)$$

eg: if you are a police officer & you want to catch criminals, you want to be sure that the person you are catching is a criminal (Precision) & you also want to catch as many criminals as possibles. (Recall).

F_1 manages this tradeoff.

$$F_\beta = (1+\beta^2)\frac{(\text{Precision} \cdot \text{Recall})}{(\beta^2 \cdot \text{Precision}) + \text{recall}}$$

(Not equal weigtage).

(v) **Log loss / binary cross entropy:**

$$L_{BCE} = -\left[y \log p + (1-y)\log(1-p)\right]$$

$p \rightarrow$ prob. of predicting 1.

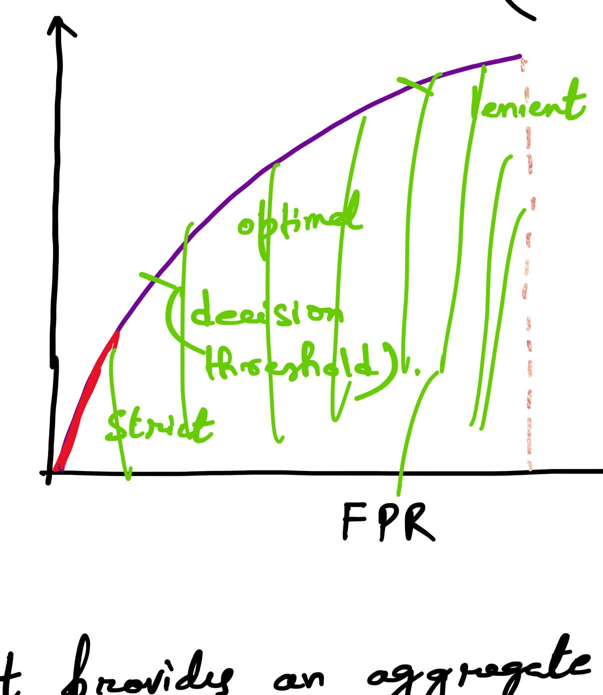
categorical CE = ?

(vi) **ROC Curve:** Sensitivity / TPR Vs FPR.

(True +ve rate)

$$TPR = Recall = \frac{TP}{(TP+FN)}$$

$$1 - \text{specificity} = FPR = \frac{FP}{(TN+FP)}$$



Receiver operating characteristic curve.

- it plots performance of classification model at all level of threshold.

AUC: it provides an aggregate measure of performance across all possible classification threshold.

Prob. of ranking a +ve examples more highly than a −ve examples.

100% wrong → AUC = 0

100% correct → AUC = 1

scale-invariant.

(measures how well prediction are ranked rather than their absolute value).