Confusion Matrix

		Pred	
		Sick Spam	Healthy Ham
Actual	Sick Spam	TP	FN Type 2
	Healthy Ham	FP Type-1	TN

Type-1 Error → Pred +ve but false

* Type-2 Error → Pred -ve, but +ve

Medical School

Pred -ve Carefully

Type-2 Critical

Spam Detector

Pred +ve Carefully

Type-1 Critical

$$\text{Accuracy} = \frac{TP + TN}{T}$$

Precision

is about Positive Pred

How many positive from total Pred +ve

$$\text{Precision} = \frac{TP}{TP + FP}$$

If Model says Positive then better +ve

Recall

ability

How many +ve detected from Actual Positiv

$$= \frac{TP}{TP + FN}$$

Precision is proportion of data points that model says relevant are actually relevant

Recall is ability to find all relevant instances

Medical School

Find all relevant TP

and reduce FN

⇒ Recall

Spam Detector

If model says posit the TP

reduce FP

⇒ Precision

Combining

In medical school, in pre-exam

for follow-up examination

If follow-up exam cost is high

⇒ High Precision

If follow-up exam cost is low

⇒ Low Precision

F1 Score or Harmonic Mean

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

HM is better since it punishes extreme values

However, F1 gives equal weight to Precision and Recall

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

$\beta = 0 \Rightarrow F_0 =$ Precision [Spa

$\beta = 1 \Rightarrow F_1 =$ HM

As β increas $F_\beta \to$ Recall

$0 < \beta < 1 \Rightarrow$ Precision

$1 < \beta < +\infty \Rightarrow$ Recall

If model has high precision $\Rightarrow$
model give less irrelevant results

If model has high recall $\Rightarrow$

model returned most of relevant
results

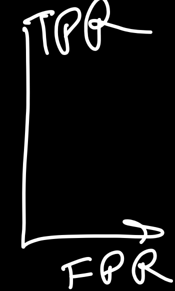
ROC Curve

Receiver Operating Characteristic

Threshold

ROC Curve shows how the recall vs precision
relationship changes as we vary threshold

By adjusting threshold value
we can have right balance



In ROC we map
False Positive Rate vs True Positive Rate

$$FPR = \frac{FP}{FP + TN} \qquad TPR = \frac{TP}{TP + FN}$$

If threshold is 1
$\Rightarrow$ +ve when $> t = 1$
$\Rightarrow$ None
$\Rightarrow$ positive $= 0 \Rightarrow TP = 0$, $FP = 0$
$\Rightarrow FPR = TPR = 0$
origin

If thresh is 0

all +ve

$\Rightarrow$ None Negn

Neg = 0 $\Rightarrow$ TN = 0 FN = 0

$\Rightarrow$

FPR = 1 TPR = 1

$\Rightarrow$ other corner

Metric is area under the curve

Higher is better

Regression Metrics

Mean Absolute Error

$$MAE = \frac{1}{n} \sum |y - \hat{y}|$$

Not Differentiable so Gradient Descent cannot be used

$f(x) = |x|$ is not differentiable at $x = 0$

$$f'(0) = \lim_{h \to 0} \frac{|0+h| - |0|}{h} = \lim_{h \to 0} \frac{|h|}{h}$$

n

$LHL = -1 \qquad RHL = 1$

A function is differentiable $\Rightarrow$
when we zoom in it looks like
straight line

Thus we use MSE

Mean Squared Error

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2$$

R2 Score

Compare the model with simplest
model

$$R2 = 1 - \frac{E_1}{E_2}$$

$E_1 \Rightarrow$ our model $\qquad E_2$ is simple model

$E_2 \rightarrow$ avg of all errors

If model is good $\Rightarrow$ less Error

$\Rightarrow E_1 < E_2$

$$\Rightarrow \frac{E_1}{E_2} \to 0 \quad \Rightarrow R_2 \to 1$$

If bad

$$E_1 > E_2$$

$$\Rightarrow \frac{E_1}{E_2} \to 1 \quad \Rightarrow R_2 \to 0$$

Bias

- occurs when algo has limited flexibility to learn true signal
- High Bias algo may miss relevant details

→ Underfitting

→ Not perform good on Training Data

→ High Bias ⇒ Less Accurate

→ Parametric Models have high bias
eg Linear Regression, LDA,
Logistic Regression

Variance

- Algorithm sensitivity to specific sets
- Occurs when limited flexibility

→ It is error from sesitivity to small fluctuations in training set

→ Overfitting

→ May fit noise

If We have High Variance

→ Get More Training Data

→ Less Features (since overfitting)

→ Increase Lambda i.. increase regularization

If High Bias

→ Increase Complexity of model

→ Get additional features

→ Decrease Lambda, less Regularizatn

Linear ML Algo

⇒ High Bias Low Variance
Under

Non-Linear ML Algo

⇒ Low Bias High Varia
Overfit

Linear Regression, Logistic Regression, LDA
⇒ High Bias, Low Variance

Decision Trees, KNN, SVM
⇒ Low Bias, High Variance

Configurations

k-nearest neighbor has low bias/high Varian
→ increase k to increase bias

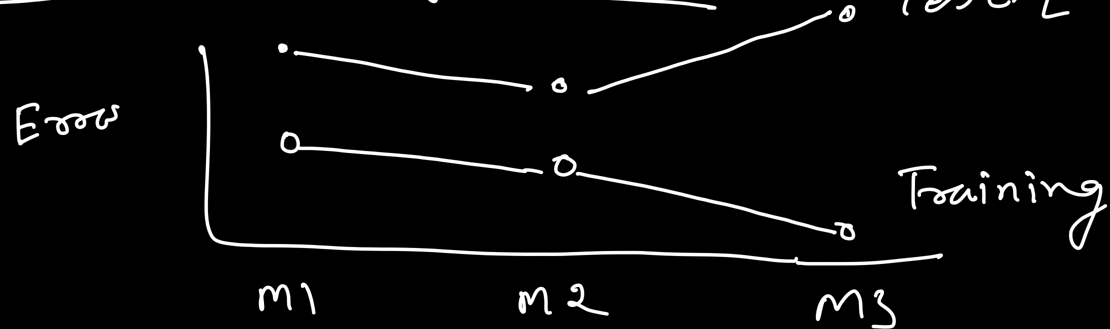
SVM has low bias, high variance
increase C-parameter that influences number of violations of margin allowed ⇒ increase bias

Linear Reg

High Bias, Low Var

reduce bias
by adding poly. fit

Model Complexity Graph



M1 ⇒ Underfitting

M2 ⇒ Right

M3 ⇒ Overfitting

To decide which model is good,
we need more data ⇒ Cross Validation

Training — Testing

Training → Training, Cross-Validation

Testing → Testing

K-Fold Cross Validation

Create K-buckets of training data and train the model k-times and each time using different bucket.

Average the result to get final

TD1 , CV1 ⇒ m1

TD — TD2 , CV2 ⇒ m2

TD3 , CV3 ⇒ m3

ENSEMBLE

→ Join Different Models to get best output

Two Approaches

Bagging or Bootstrap Aggregation

Boosting

Bagging < Avg
Voting

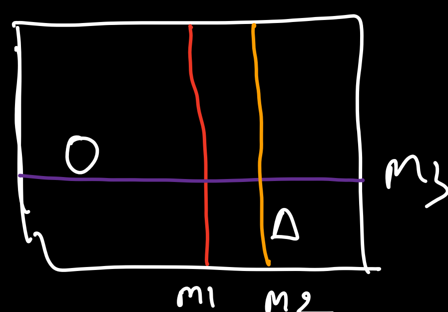
Boosting → combine based on model's strength.

M1 for class1
M2 for class2

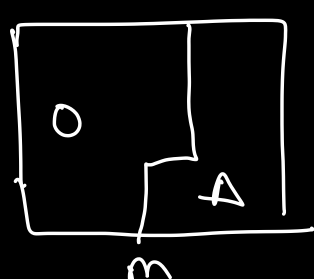
Bagging → Avg or Voting

In Voting

Weak Learners ⇒ Strong Learners



Ensemble



AdaBoost

many implementatn

Boosting $\Rightarrow$ strength of one model

2^{nd} Learner classify the miss-classified points of 1^{st} Weak Learner

3^{rd} Learner classify the mis-classified points of 2^{nd} Learner

How it happens

m1 $\rightarrow$ equal weights say 7 Correct, 3 Incorrect

Increase weights of mis-classifn pos

$$m1 \begin{cases} 1+1+1+1+1+1+1 & = 7 \\ \frac{7}{3}+\frac{7}{3}+\frac{7}{3} & = 7 \end{cases}$$

$$m2 \begin{cases} \text{Corre } 11 \\ \text{Incorr } 3 \end{cases}$$

According to weigh

$$m2 \begin{cases} 11 & = 11 \\ \frac{11}{3}+\frac{11}{3}+\frac{11}{3} & = 11 \end{cases}$$

M3 — 19

M3 — 3

continue or stop and Combine

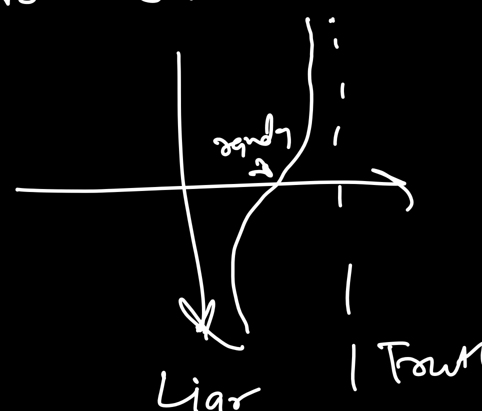
<u>Combine</u>

Large +ve weight to Truth Model

Large −ve weight to Liar Model

Zero weight to random model

$$y = \ln\left(\frac{x}{1-x}\right)$$



$$\lim_{x \to 0^+} y = -\infty \quad \Big| \ \ln 0$$

$$\lim_{x \to 1} y = \infty \quad \Big| \ \ln \infty$$

$$\lim_{x \to 0.5} y = 0 \quad \Big| \ \ln 1$$

$$\text{Weight} = \ln\left(\frac{\text{Correct}}{1-\text{Correct}}\right)$$

$$= \ln\left(\frac{\text{Correct}}{\text{Incorrect}}\right)$$

<u>Combine</u>

	m1	m2	m3
Correct	7	11	19
Incorrect	3	3	3
	$\ln\left(\frac{7}{3}\right)$	$\ln\left(\frac{11}{3}\right)$	$\ln\left(\frac{19}{3}\right)$
	0.84	1.3	1.84

m1

0.84	-0.84

m2

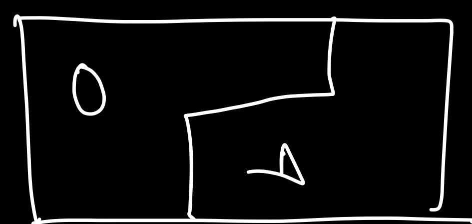
1.3
-1.3

m3

1.84	-1.84

0.84 1.3 1.84	-0.84 1.3 1.84	-0.84 1.3 -1.84
0.84 -1.3 1.84	-0.84 -1.3 1.84	-0.84 -1.3 -1.84

3.98	2.3	-1.38	+ve cl1
1.38	-0.3	-3.98	-ve cl2



⇒ Ada Boost