~~Week~~ L2 (Improving building strategies)

Orgthogonalization

is a process in linear algebra where a set of vectors is transformed into a new set of vectors that are mutually perpendicular and optionally have a unit length.

e.g.

Car has Brake, steering, Accelerator.
If we create a joy stick performing all function ~~we~~ ~~w~~ $\underbrace{\text{joy stick}}$ that's orthongonalization is.

Classifier	Perecision	Recall	F1 Score
A	95%	90%	92.4%
B	98%	85	91.0%

∴ ~~Precision how % model detect cats~~
∴ Precision: how many cats are present in data
∴ ~~Are~~ Recall: How many % of cats are detection.
~~Recall~~ F1 Score: averge of precision and cat or harmonic precision

F1 Score is greater in classifier A so we choose A.

Satisficing and Optimizing metric.

let we want accuracy and running performance.

Classifier	Accuracy	Running time
A	90%	80ms
B	92%	95ms
C	95%	1500ms

cost = accuracy - 0.5 x running time

running should be ≤ 100 ms

When we want to trigger word like in Alexa, we want running time and accuracy

Optimizin (Accuracy) ~~an~~

	Accuracy	Run time
A	89%	80 ~~100ms~~ ms
B	87%	~~70ms~~ 95ms

Optimizing, we choose 89%
we will chose ~~B~~ A

~~Statis~~ Satisfying (latency)

we will choose model B as a satisfying model bcz we want accuracy and running time both

---

Distribution of dev sets

Dev	Test
India	Nigeria
China	Thailand
Usa	UK
France	

X

never ever

If get data from those countries we will their dev and test set.

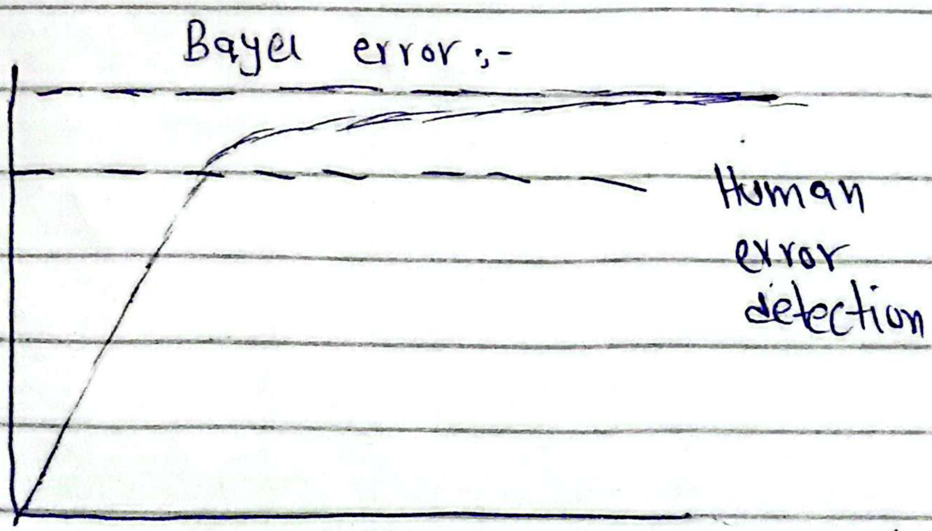
L# When to change dev and teste sete.

Suppose our model on high-resolution picture detection is

A	95%
B	93% — Chosed

But B also performs good on blurry images as compared A then we will chose B model.

T# Human level Performance.



When machines ability surpass human level performances and after that they reach a limit and cannot surpass this this

How to choose human error?
like a senior doctor has 0.5% chance of detecting image then "human error" will be "0.5%" optimal point is called "bayes error".

I # Avoiding Bias.

	Human	1%	7.5
machine	training	8%	8%
	Dev	10%	10%

variance

focus on bias

⟹ when difference b/w human and machine detection ~~we~~ is very high we will go with reducing bias.

⟹ When difference b/w human and machis is no very high $\boxed{7.5}$ & $\boxed{8\%}$ we will decrease variance between machine dev & traning error.

I # Supassing Human level Performance.

In natural process surpassing humans are defficult to surpass, I mean detection of images, vision recognition. but supass humans in databases things, Like many things

Like voice detection, noisy voices which are defficult for humans machine can surpass them/ML mode can supass them -

IMPRoving Model Performance.

If u want to ~~I~~ improve, improve

Human level error

make leu bias

compare

Training Error

variance.

Deve error

Module 2:

Carrying out Error analysis.

Incorrectly labeled examples

x 

y labels labels labels

note: When a model makes a prediction, it outputs probabilities for each class.

Class	probability (detecting dog pic)
cat	0.05
Dog	0.94
Airplane	0.01

We will compare labels by actual label to check mislabeled.

- Build dev/set
- Build initial system
- Use variance/Bias analysis & Error and then Iterate.

Don't need ~~to~~ try create complex model first make an easy one the move on.

Traing a set

250,000 images are high quality.
2500 are low quality

If we shuffle and create dev/test set

T.S (high quality+low)	D.S	T.S

This model will work best on high quality pictures but number of pictures are very low so it will priotize high quality picture to detect cats.

Speech recognition

Bias Variance Data Distribution

error

Train =	1%	
dev =	1.5%	this is data mismatched
Test =	10%	problem

Trainning	10%	10%	variance small
Training deve error	11%	11%	
Dev error	11%	20%	Data mismatched error
High bias error		~~High variance~~	

Human level	4%	Bias
Training error	7%	variance
Training dev set	10%	
Dev erro	12	dat mismatched

Addressing Data Mismatch

When transfer learning makes sense

⇒ Task A and B have same input x.
⇒ You have a lot more data for task A than task B.
⇒ low level features could be helpful for learning B.

Defination

where a model trained on one task is reused on a related task.

example

* Hospital want to build ai system ~~but they~~ to ~~have~~ detect lung cancer from chest X-ray.
* They have a very limited data
* Training a deep NN like a CNN requires millions of pictures.

Multi-task example

Simplified autonomous driving example

Padestrian 0

Cars 1

Stop signs 1

Traffic lights 0

⋮

Model is solving 4 problems so it is multitasking learning.

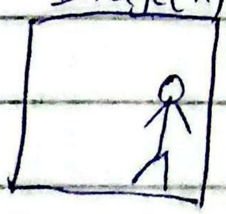
When to do this:

Where small thing effect predictions like in Simplified autonomous driving example.

End to End learning

Speech recognition example

Image(x)



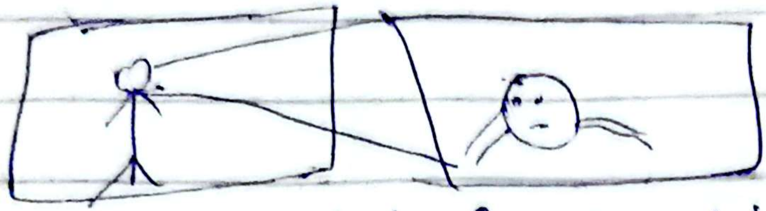
→ (y) Identity

this not we like that

get image → find identity

~~first~~ face is captured ⇒

1) first capture picture



ii) then detect face by detecting

iii) Identify persons id.

end to end approach

Use pipline.

When to end to end learning:-

Pros:-

Allow the data to speak

But require large data like for detected X-ray pictures to detect childs hand.

e.g

image ⟶ steering

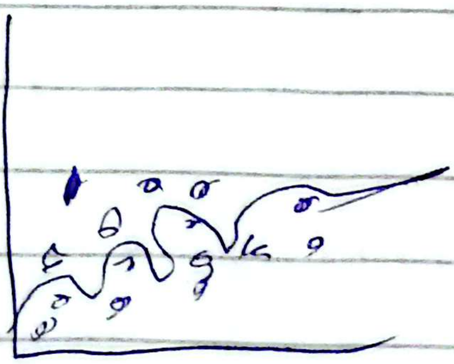
this is not a good approach.

MIT Class 1

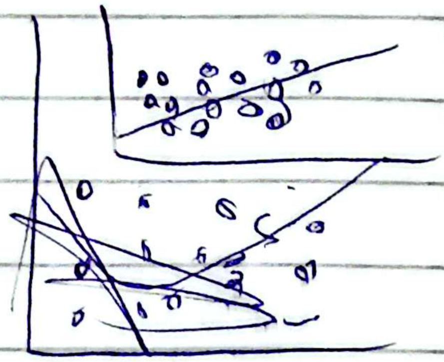
process of solving descent gradient is backpropagation.

How to choose α (Learning rates)
use different learning rates

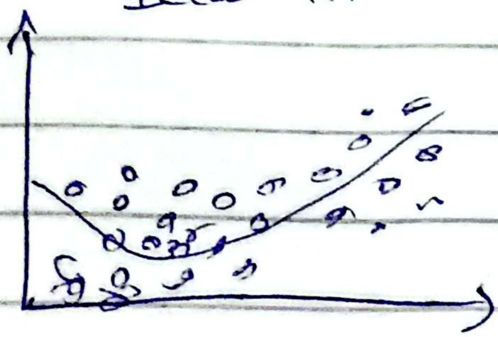
overfitting



underfitting



Ideal fit



Regularization Early stopping