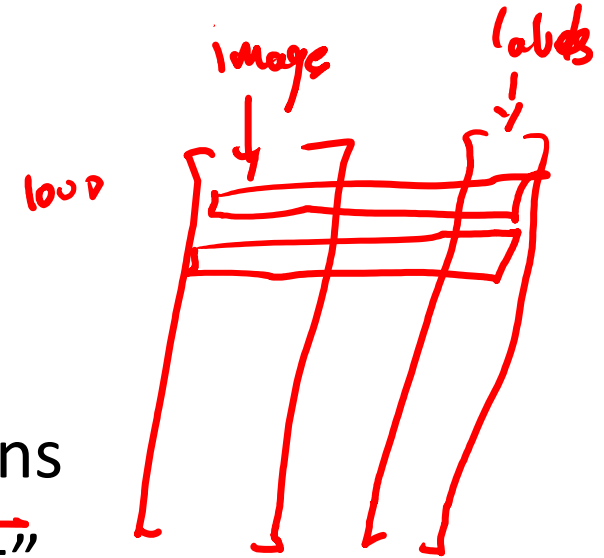**Georgia Tech**

Error Measures and Testing

Nakul Gopalan

Georgia Tech

Best Practices

- Train vs Test data: Do not mix them!!
- K-fold validation for hyper-parameter optimizations
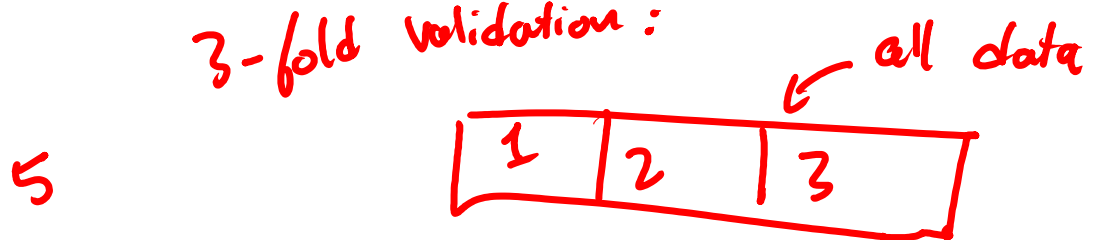- Older books would suggest using a "validation set"

image   labels

loop

3-fold validation:

all data

| 1 | 2 | 3 |

5

train   test

1,2   3

2,3   1

1,3   2

$\sum_{i \in 1,3} test(i) = test\ statistic$

80 % train
20% test

train	valid.	test
60%	20%	20%

Error measure

- True classifier -> f(x)
- Learned classifier, based on a hypothesis -> h(x)
- Error: $E(h, f)$
- Almost always pointwise definition: $e(h(x), f(x))$
- Examples:
  - Squared error: $e(h(x), f(x)) = (h(x) - f(x))^2$
  - Binary error: $e(h(x), f(x)) = ||h(x) \neq f(x)||$

From pointwise to overall

Overall error = E(h, f) = average of pointwise errors $= e\big(h(x) - f(x)\big)$

In-sample error: $E_{in}(h) = \sum_1^N e(h(xn) - f(xn))$

training set size

train

Out-of-sample error: $E_{in}(h) = \mathbb{E}[e\big(h(x_n) - f(x_n)\big)]$
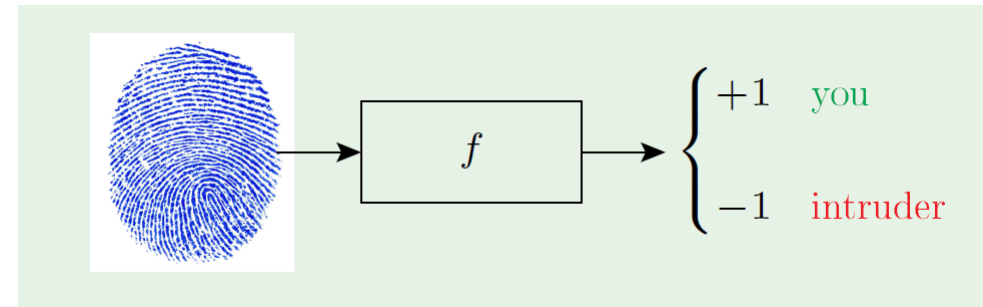
out

test

How to choose the error measure

Finger-print verification:

Two Types of Errors:
- False accept (False Positive)
- False reject (False Negative)

Correct answers are True Positive
and True Negatives

How do we penalize these errors??

$$\begin{cases} +1 & \text{you} \\ -1 & \text{intruder} \end{cases}$$

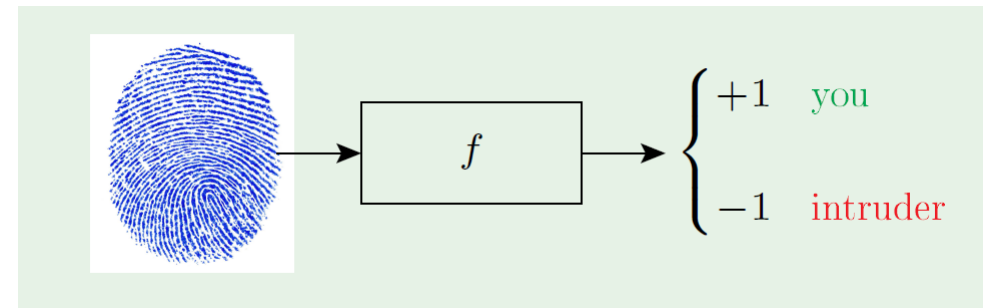|	f	
	+1	-1
h +1	no error	false accept Type 2 error
-1	false reject/ Type 1 error	no error

How to choose the error measure

Finger-print verification:

Two Types of Errors:
- False accept
- False reject

How do we penalize errors:

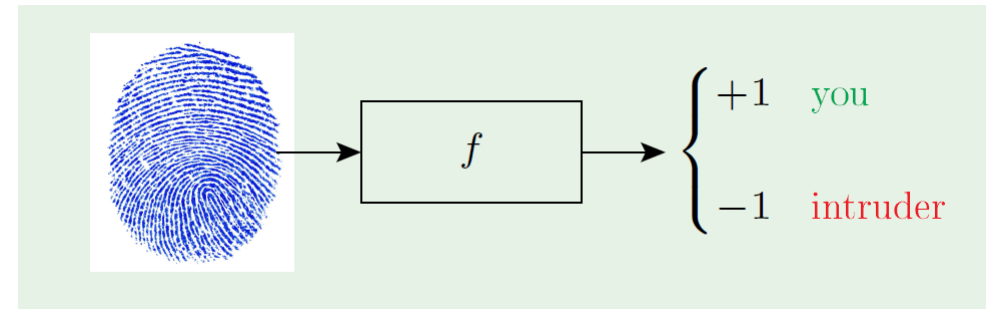|	f	
	+1	-1
h +1	0	+1
-1	+1	0

How to choose the error measure - Supermarket

Supermarket verifying customers

False reject is costly:
Real customer not let in! Customer annoyed!

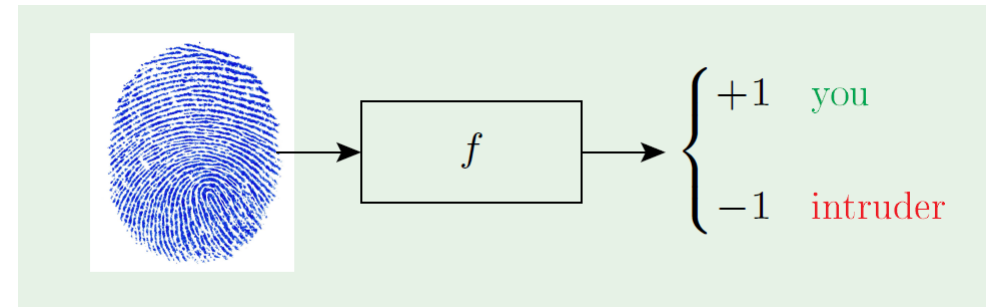False accept is minor. Some random
customer gets in; not that expensive.

		f	
		+1	-1
h	+1	0	+1
	-1	+100	0

How to choose the error measure - CIA

Finger-print verification for security

- False accept is BAD!!

- False reject is ok, try again ☺

		f	
		+1	-1
h	+1	0	+10000
	-1	+1	0

Error measure for rare data cancer detection
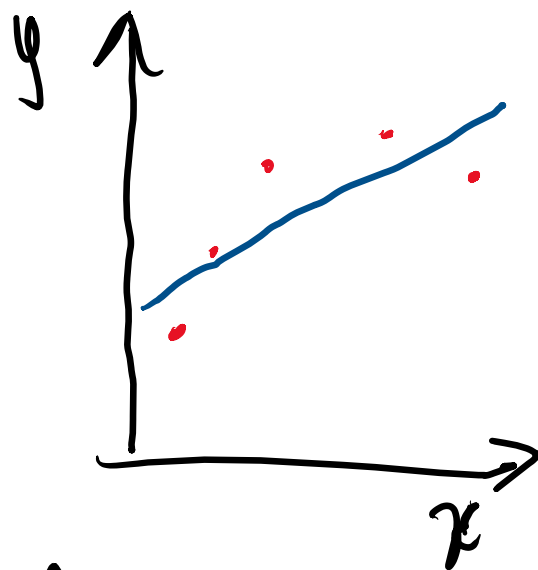
Finger-print verification for security!!

- False accept is ok.

- False reject is terrible!!

More samples for non-cancer
and less for cancer.

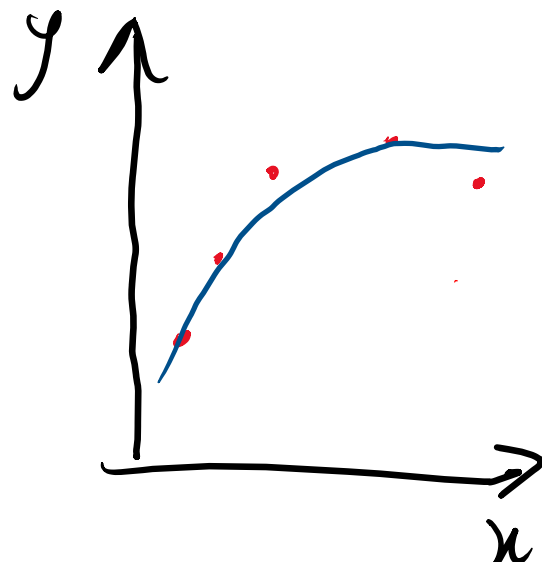Weight by data asymmetry and
the penalty of missed detection.

f

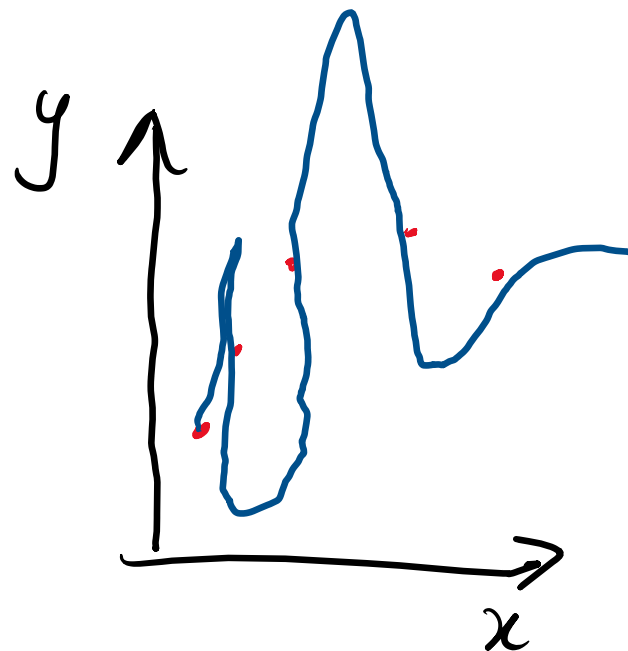|		+1	-1
h	+1	0	+1
	-1	+1 ooo	0

Bias vs Variance

$$\hat{y} = \theta_0 + \theta_1 x$$

Biased + Underfit

$$\hat{y} = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

Just right

$$\hat{y} = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots \dots \theta_{10} x^{10}$$
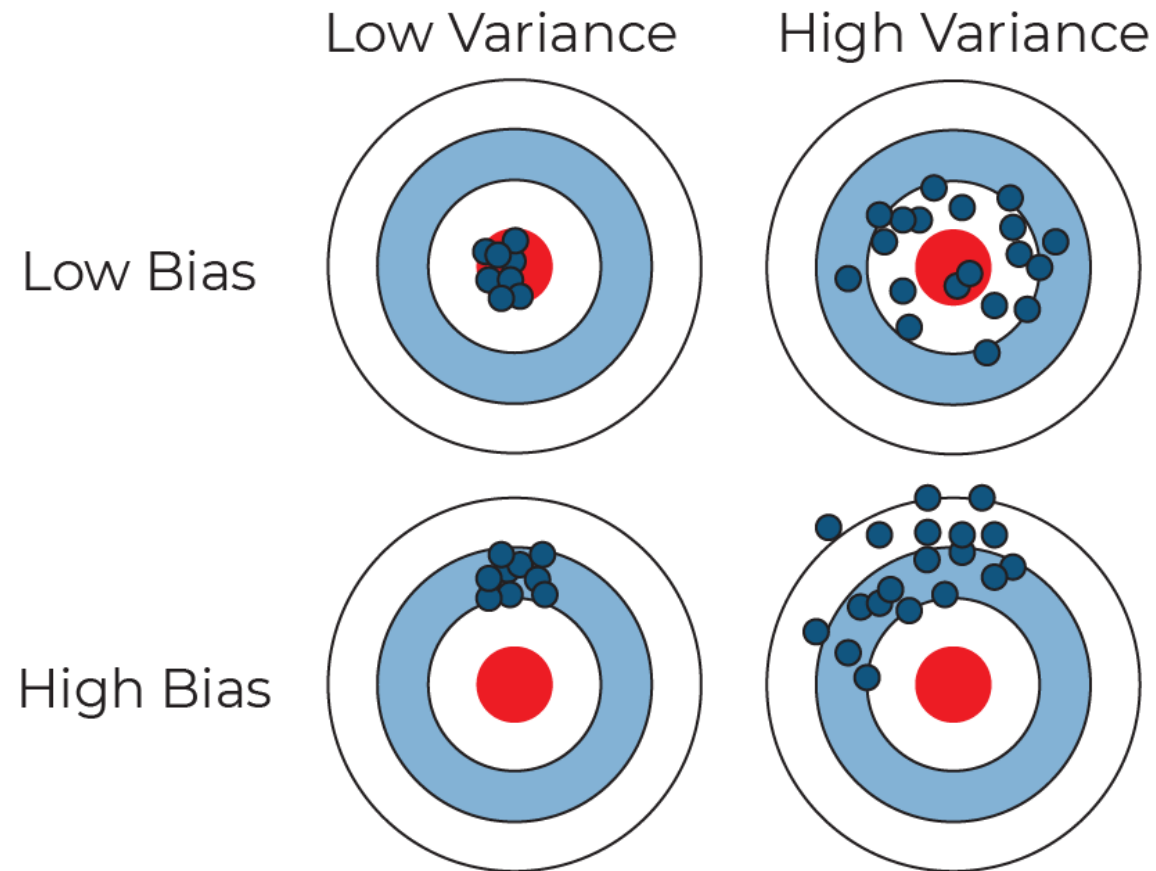
Overfit + high Variance

Bias vs Variance

Figure by Scott Fortmann-Roe