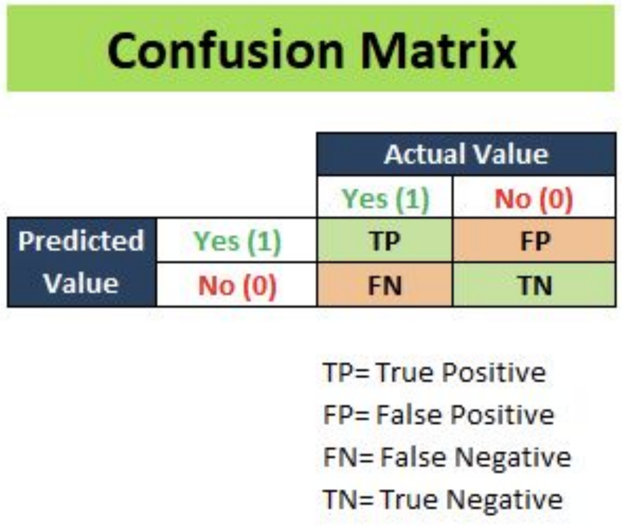
CYBER 207
Applied Machine Learning for Cybersecurity

Summer 2023

Week 5

Classification Metrics



Confusion Matrix

		Actual Value	
		Yes (1)	No (0)
Predicted Value	Yes (1)	TP	FP
	No (0)	FN	TN

TP= True Positive

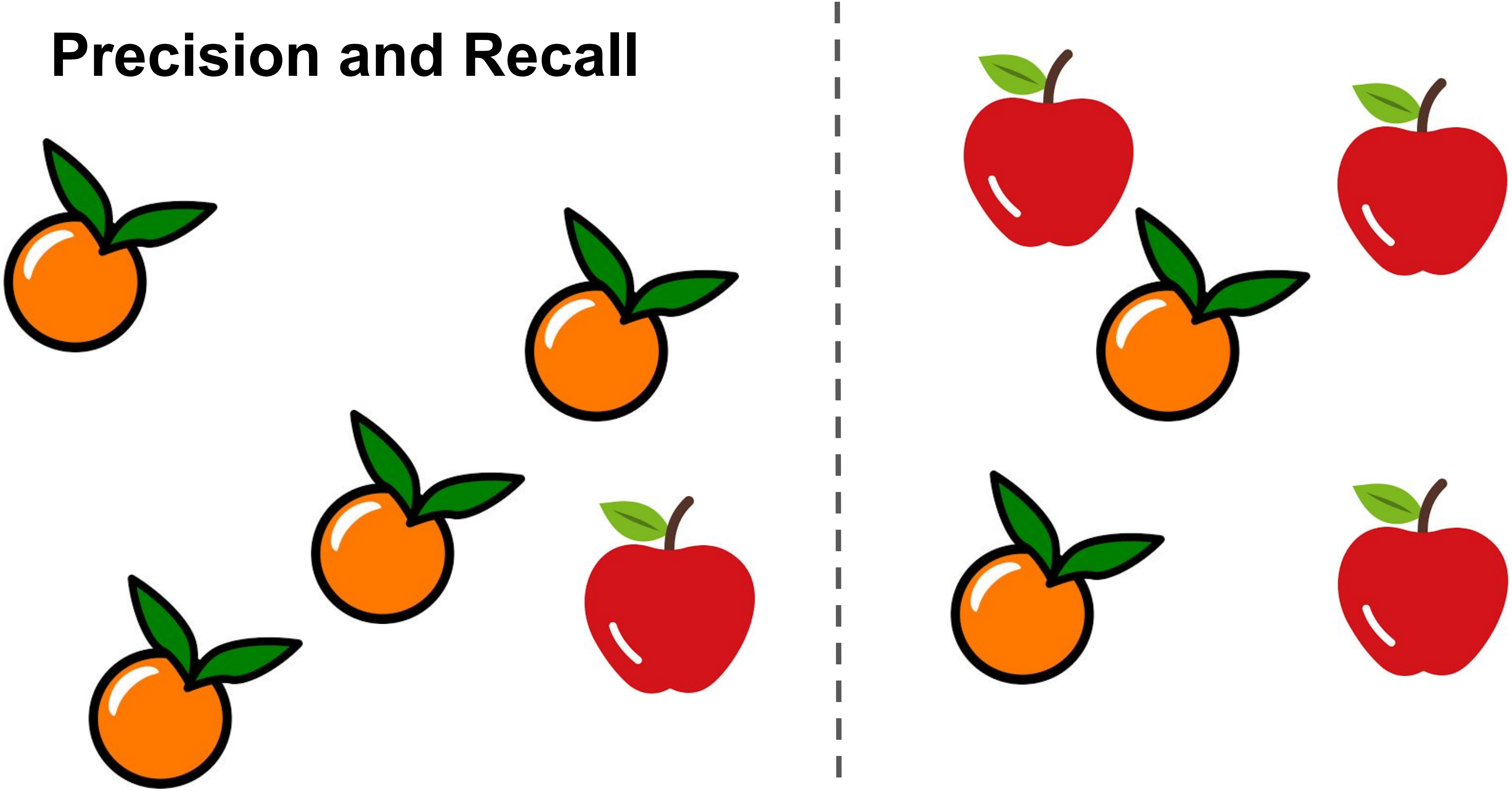
FP= False Positive

FN= False Negative

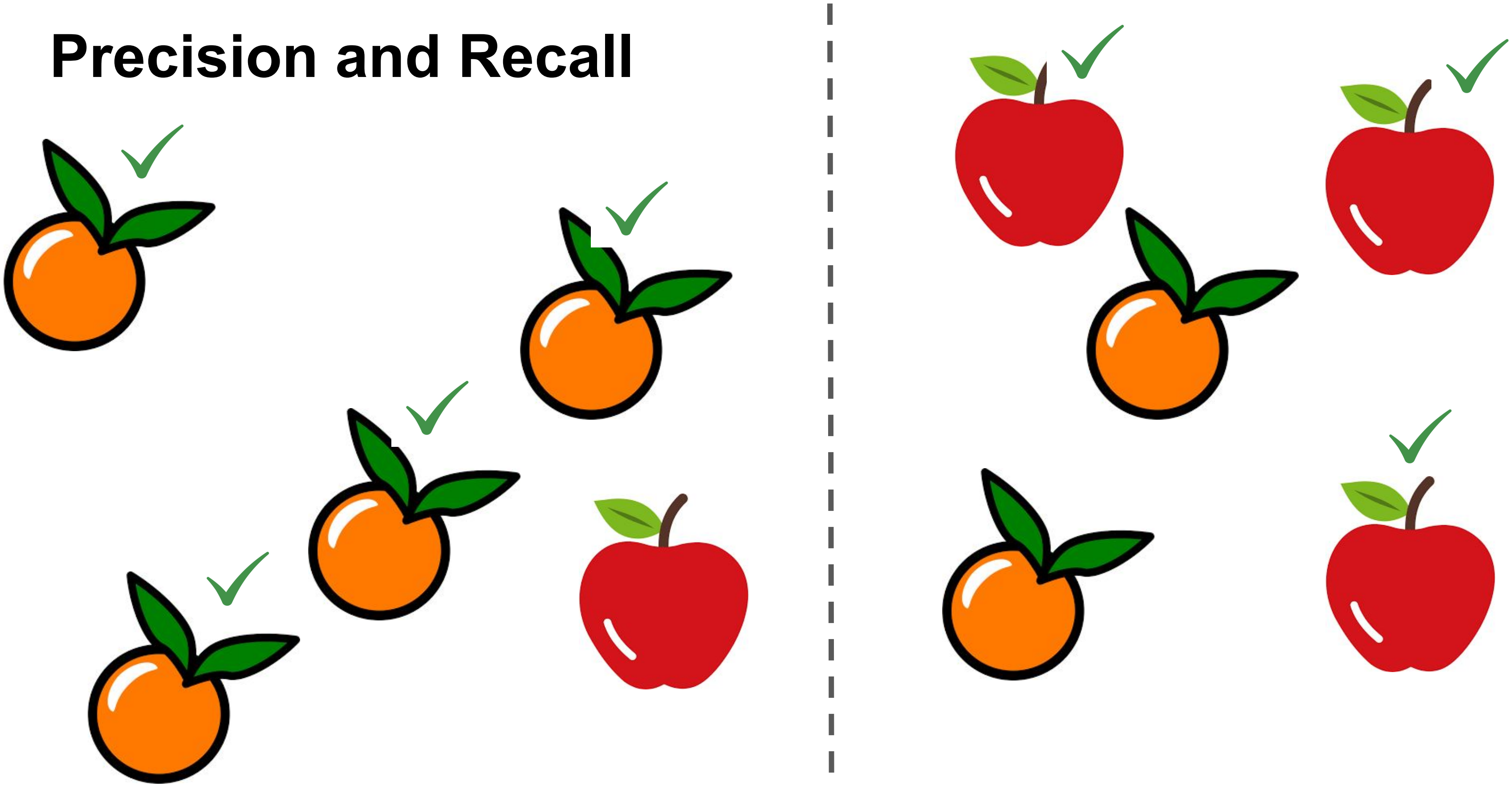
TN= True Negative

- If you have supervised data, you will want to maximize an objective function.
  - **Precision**: $TP \div (TP + FP)$ % positives correctly identifed
  - **Recall**: $TP \div (TP + FN)$ % existing positives identified
  - **Optimal point** on ROC (precision/recall) curve
  - **Accuracy**: $(TP + TN) \div (TP + TN + FP + FN)$
  - **F-test**: $2 \cdot (P \cdot R) \div (P + R)$

Precision and Recall

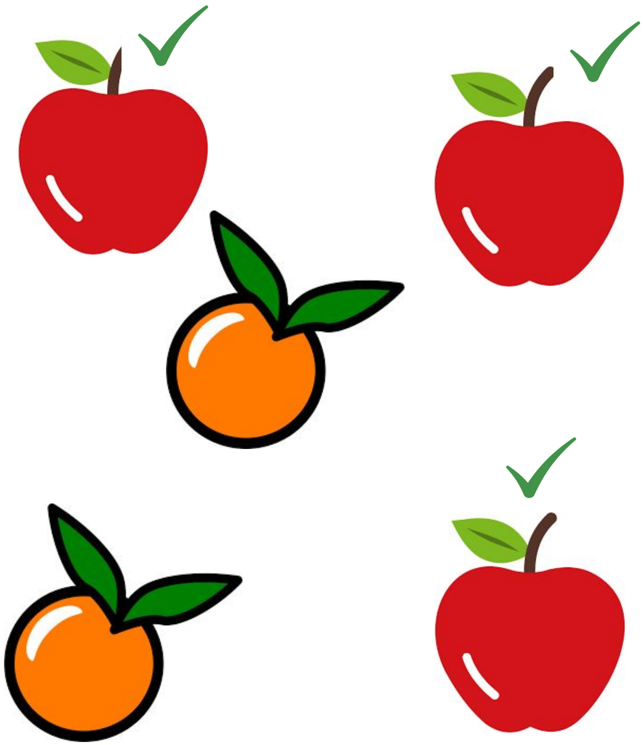


Precision and Recall



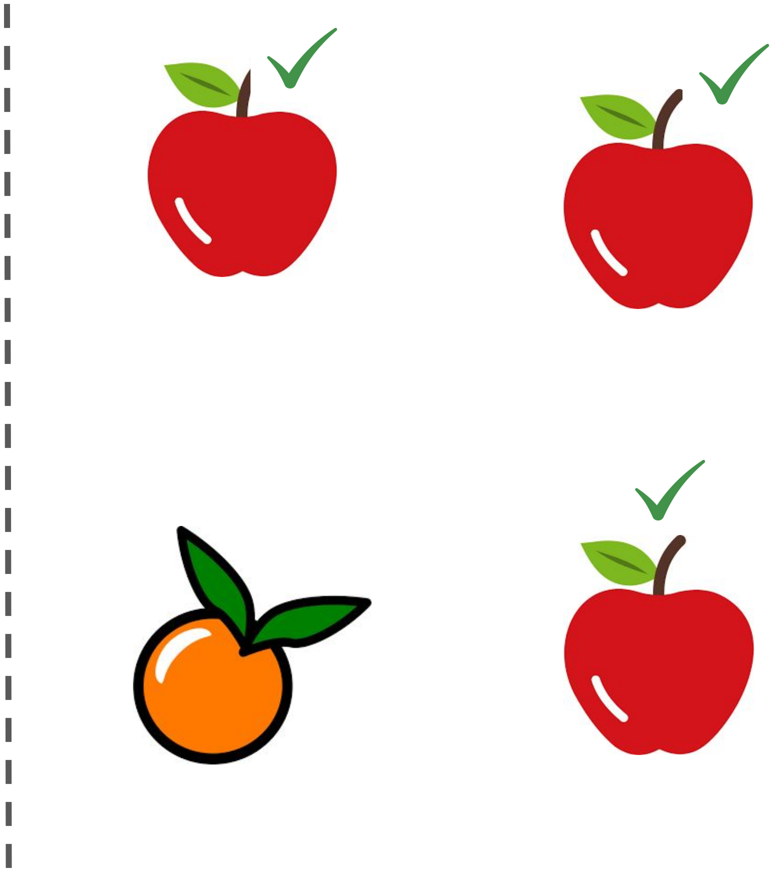
Precision

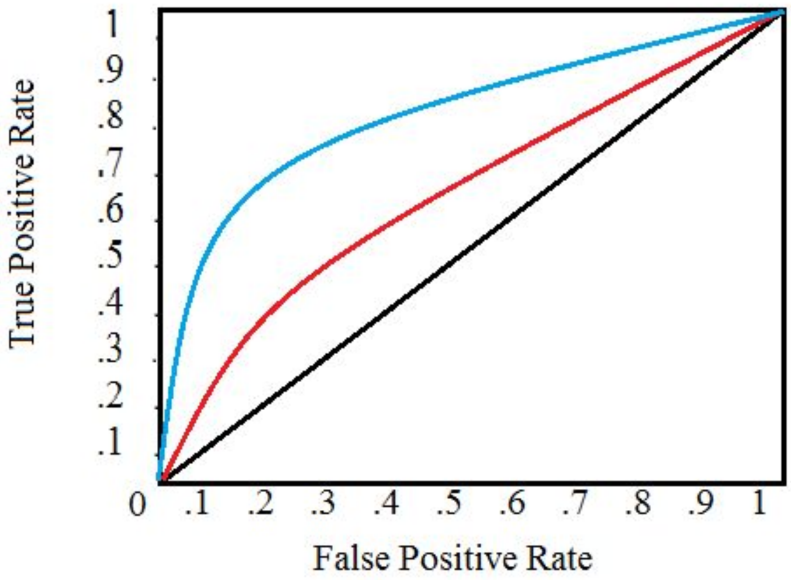
(Total apples correct)/ (Total apple side observations)
= ⅗ = 60%

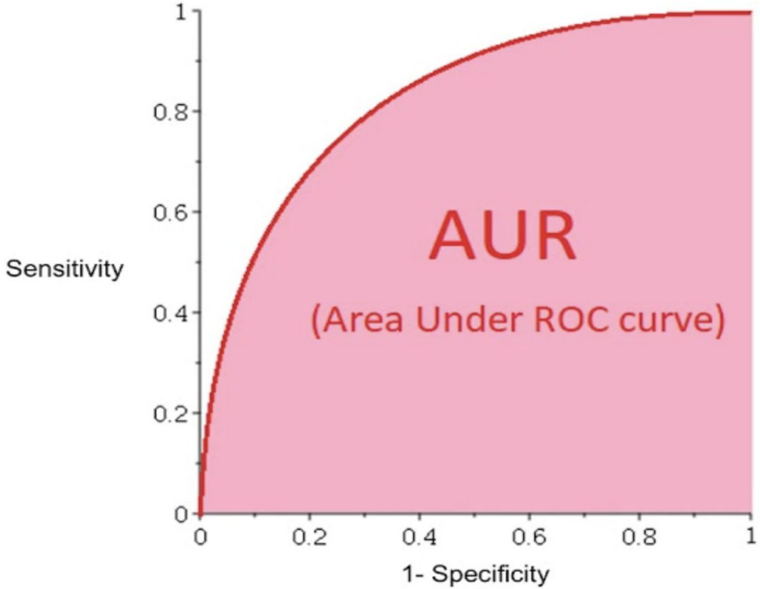


Recall

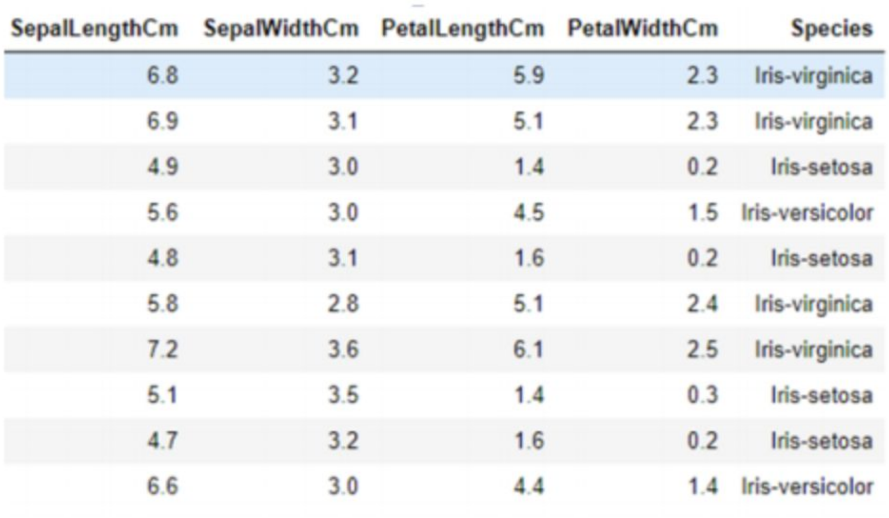
(Total number of apples correct)/ (Total actual apples)
= ¾ = 75%




True Positive Rate
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
False Positive Rate

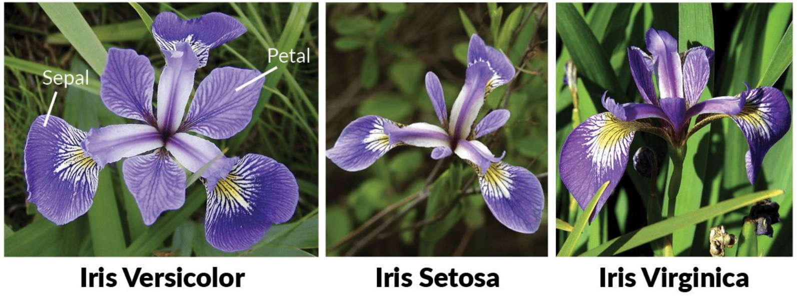

Sensitivity
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1- Specificity
AUR
(Area Under ROC curve)

Multiclass Classification

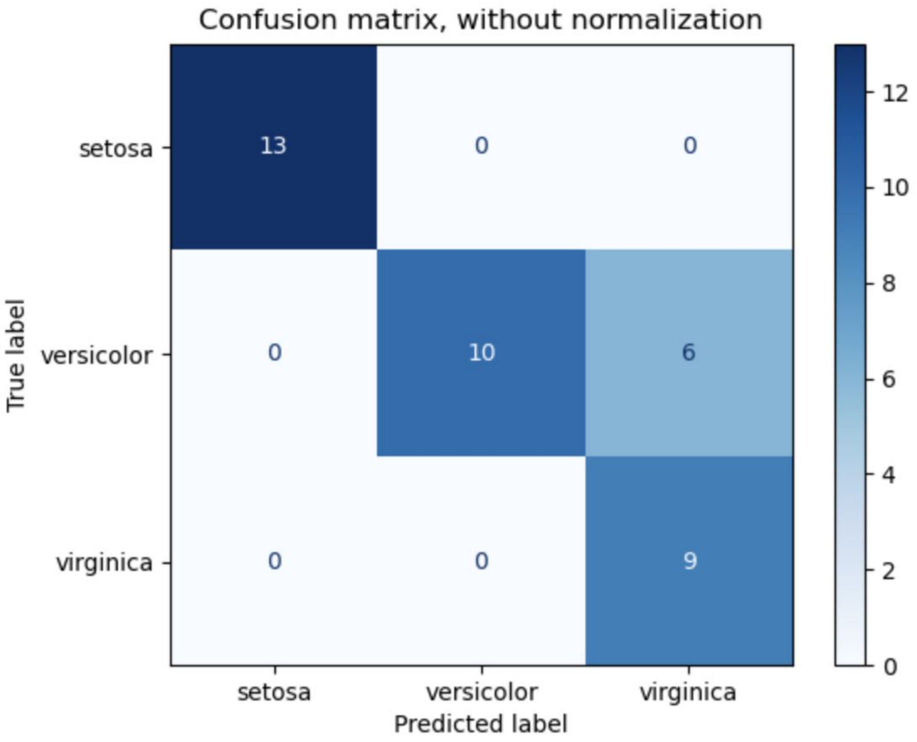


SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
6.8	3.2	5.9	2.3	Iris-virginica
6.9	3.1	5.1	2.3	Iris-virginica
4.9	3.0	1.4	0.2	Iris-setosa
5.6	3.0	4.5	1.5	Iris-versicolor
4.8	3.1	1.6	0.2	Iris-setosa
5.8	2.8	5.1	2.4	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
5.1	3.5	1.4	0.3	Iris-setosa
4.7	3.2	1.6	0.2	Iris-setosa
6.6	3.0	4.4	1.4	Iris-versicolor

Fig.1: Iris dataset having three categories

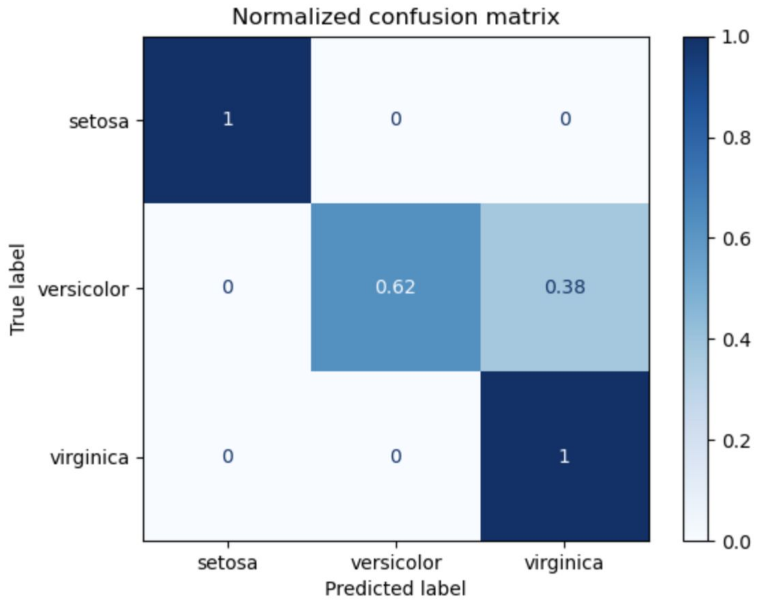


Multiclass Classification Confusion Matrix



Confusion matrix, without normalization

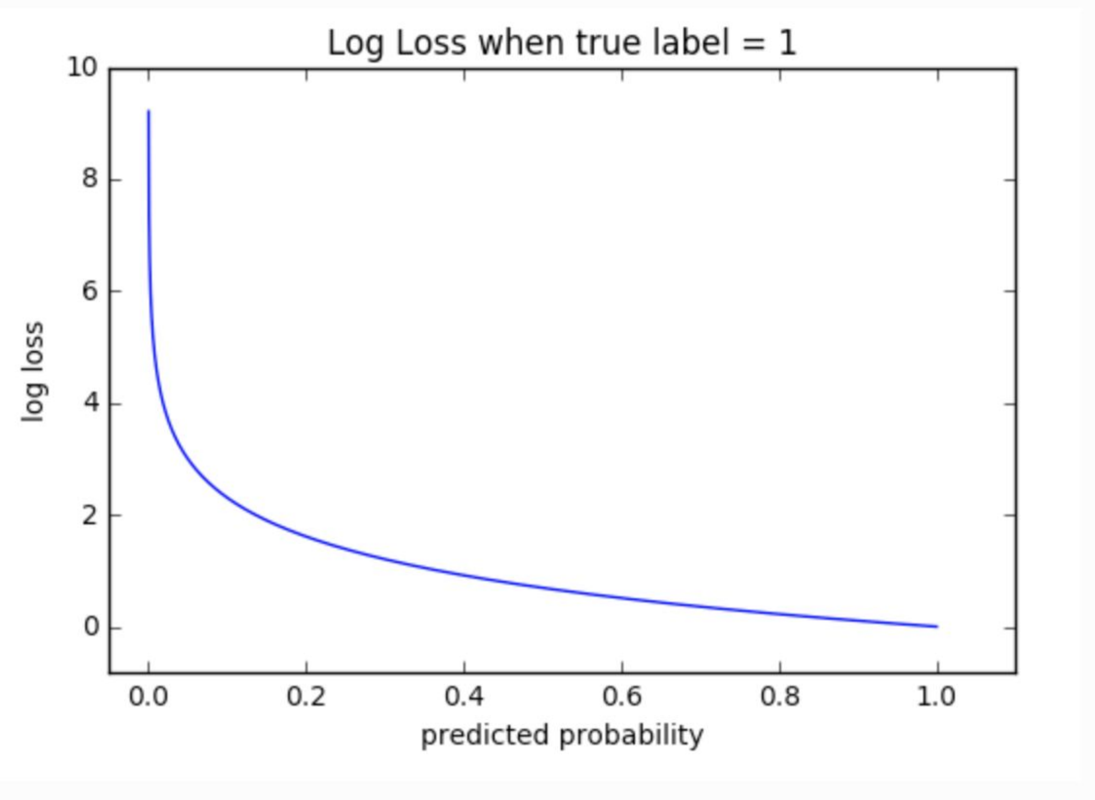
True label \ Predicted label	setosa	versicolor	virginica
setosa	13	0	0
versicolor	0	10	6
virginica	0	0	9



Normalized confusion matrix

True label \ Predicted label	setosa	versicolor	virginica
setosa	1	0	0
versicolor	0	0.62	0.38
virginica	0	0	1

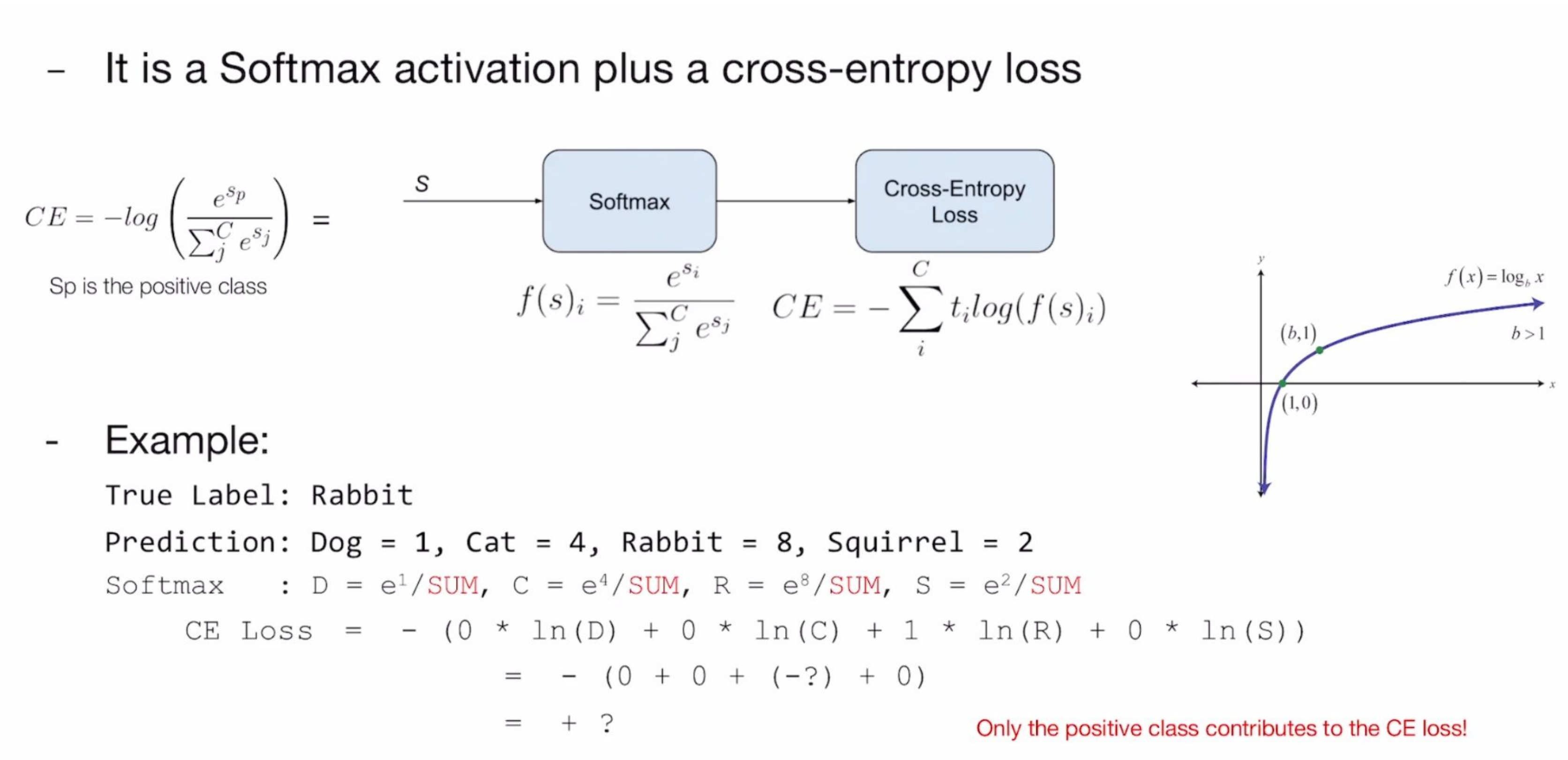
Cross Entropy



$$-(y \log(p) + (1 - y) \log(1 - p))$$

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

Categorical Cross Entropy Loss (Softmax Loss)



- It is a Softmax activation plus a cross-entropy loss

$$CE = -log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right) =$$

Sp is the positive class

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \qquad CE = -\sum_i^C t_i log(f(s)_i)$$

- Example:

True Label: Rabbit

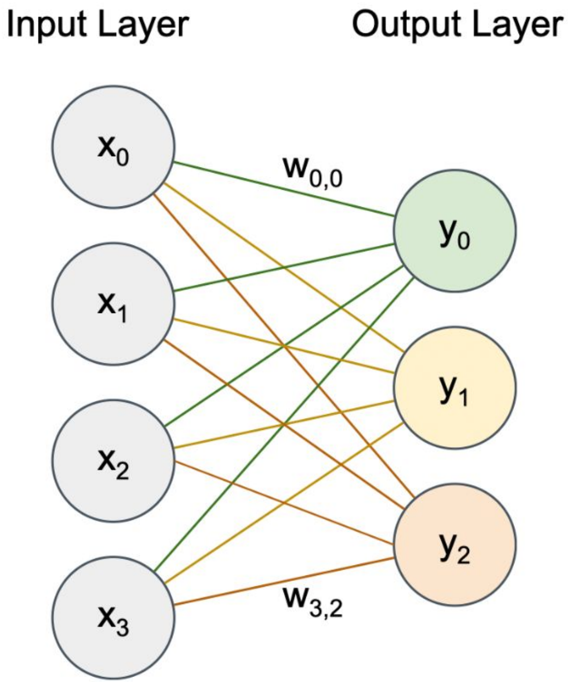
Prediction: Dog = 1, Cat = 4, Rabbit = 8, Squirrel = 2

Softmax : D = e^1/SUM, C = e^4/SUM, R = e^8/SUM, S = e^2/SUM

```
    CE Loss  =  - (0 * ln(D) + 0 * ln(C) + 1 * ln(R) + 0 * ln(S))
                          =  - (0 + 0 + (-?) + 0)
                          =  + ?
```

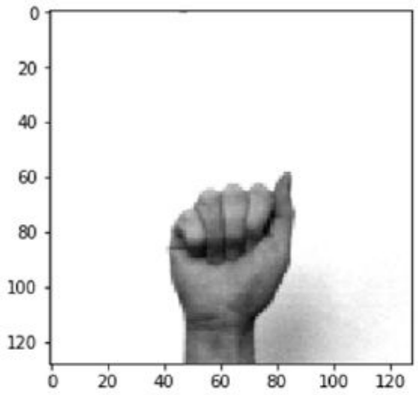
Only the positive class contributes to the CE loss!

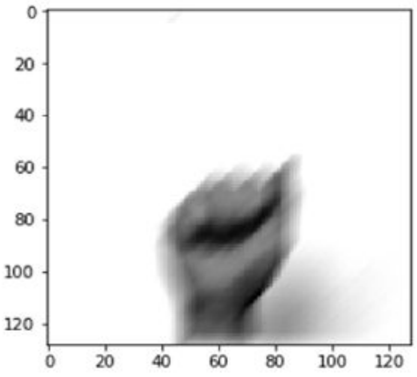
Logistic Regression Network Graph

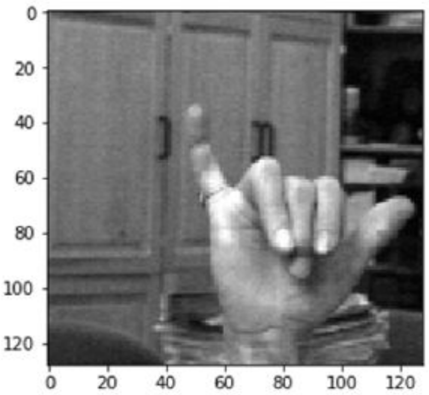


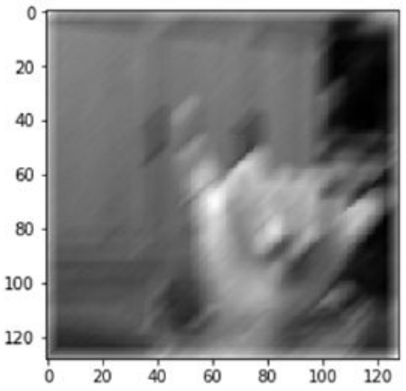
$$\begin{pmatrix} x_{0,0} & x_{0,1} & x_{0,2} & x_{0,3} \\ x_{1,0} & x_{1,1} & x_{1,2} & x_{1,3} \end{pmatrix} \begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} \\ w_{1,0} & w_{1,1} & w_{1,2} \\ w_{2,0} & w_{2,1} & w_{2,2} \\ w_{3,0} & w_{3,1} & w_{3,2} \end{pmatrix} \xrightarrow{\sigma} \begin{pmatrix} y_{0,0} & y_{0,1} & y_{0,2} \\ y_{1,0} & y_{1,1} & y_{1,2} \end{pmatrix}$$

Linear Model Limitations



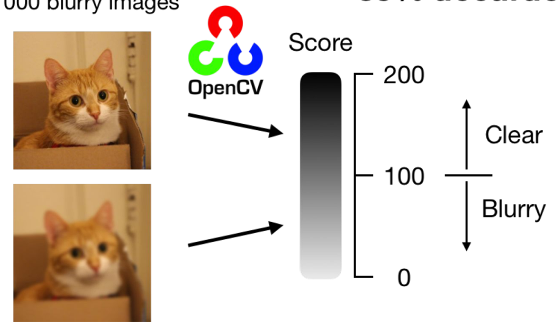




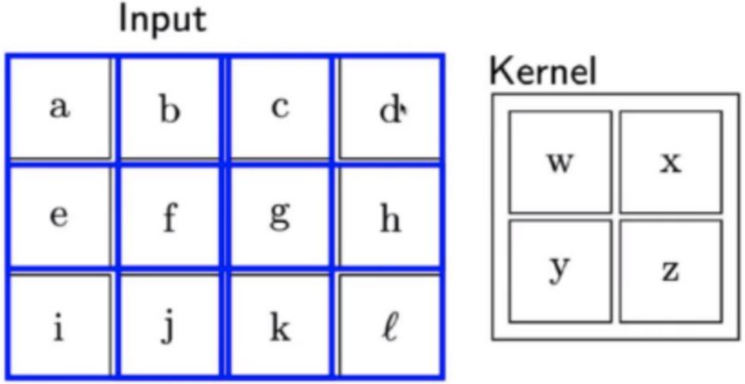


- 1000 clear images
- 1000 blurry images

85% accuracy



Convolution Operation



Input

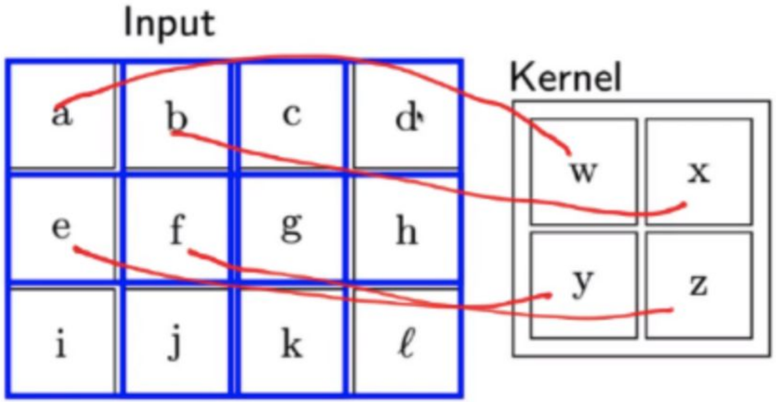
a	b	c	d
e	f	g	h
i	j	k	ℓ

Kernel

w	x
y	z

$$S_{ij} = (I * K)_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{n-1} I_{i+a,j+b} K_{a,b}$$

Convolution Operation



Input

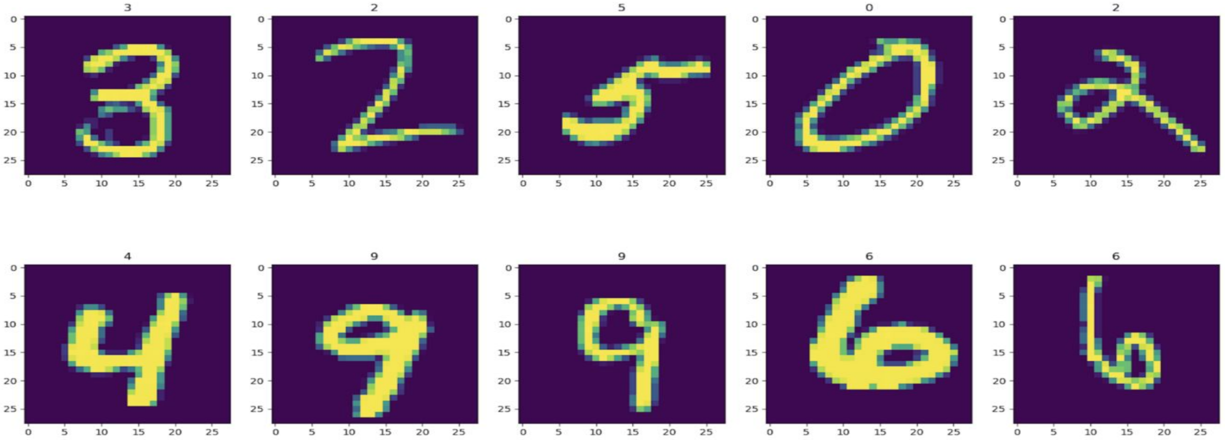
a	b	c	d
e	f	g	h
i	j	k	ℓ

Kernel

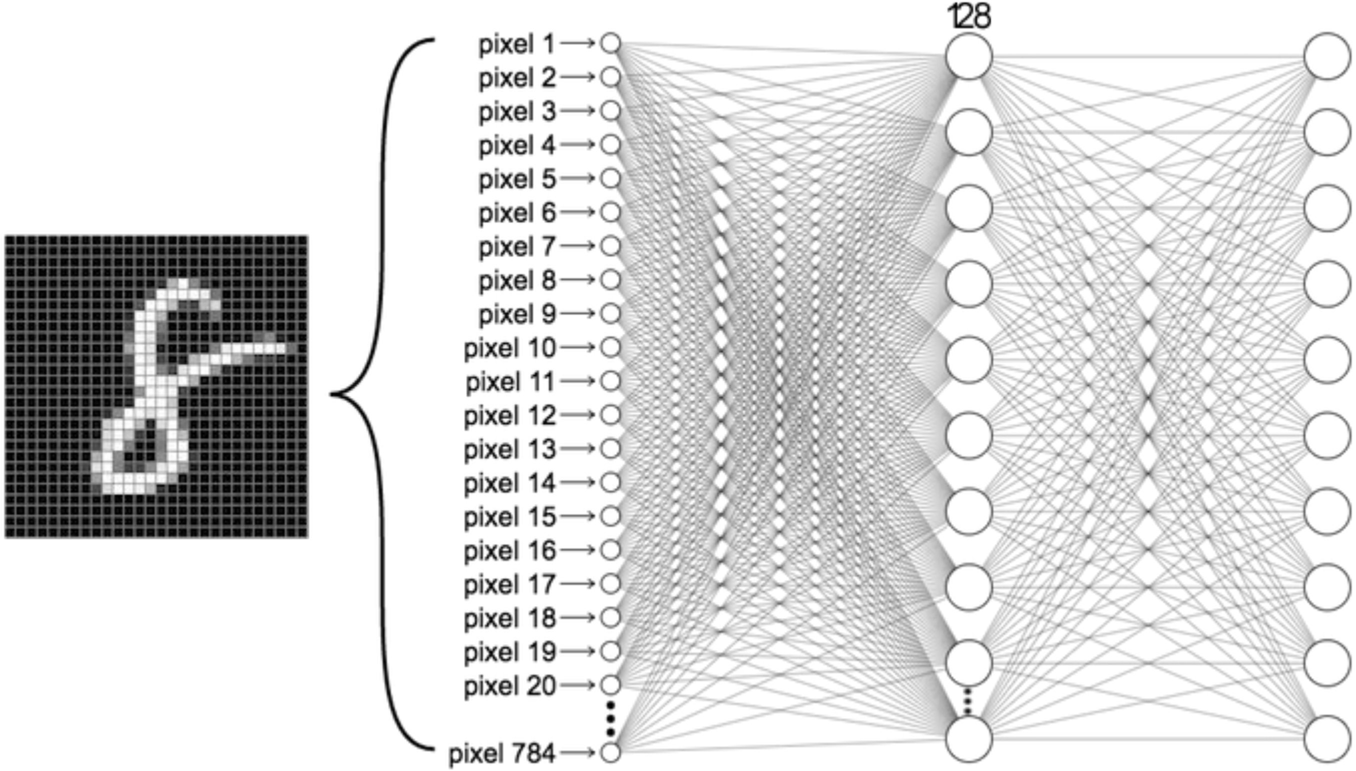
w	x
y	z

aw+bx+ey+fz	bw+cx+fy+gz	cw+dx+gy+hz

Digit Classification Problem



Digit Classification Problem



Code Review