Build It, Break It
The Language Edition

Are you a **Builder**?

Do you think your approach to natural language processing problems can **withstand attacks**? Is it **robust** to unexpected inputs, or domain mismatch? Will a **rule-based** approach hold up better than a **learning-based** approach?

Participate in our shared tasks by solving an NLP problem—by rules or by learning—and learn how your approach stacks up against others on **adversarially, linguistically constructed** inputs.

Or maybe you're a **Breaker**? Click to flip...

read shared task details

join our mailing list for announcements

join our workshop at emnlp 2017

designed & implemented by

Emily M. Bender

Hal Daumé III

Allyson Ettinger

Sentiment analysis
Logistic Regression

Logistic Regression (LR)

Logistic Regression

a neural network can be viewed as a series of logistic regression classifiers stacked on top of each other.

Generative vs. Discriminative

NB is a Generative Model

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{\sum_{H \in \mathcal{H}} P(E|H)\,P(H)}$$

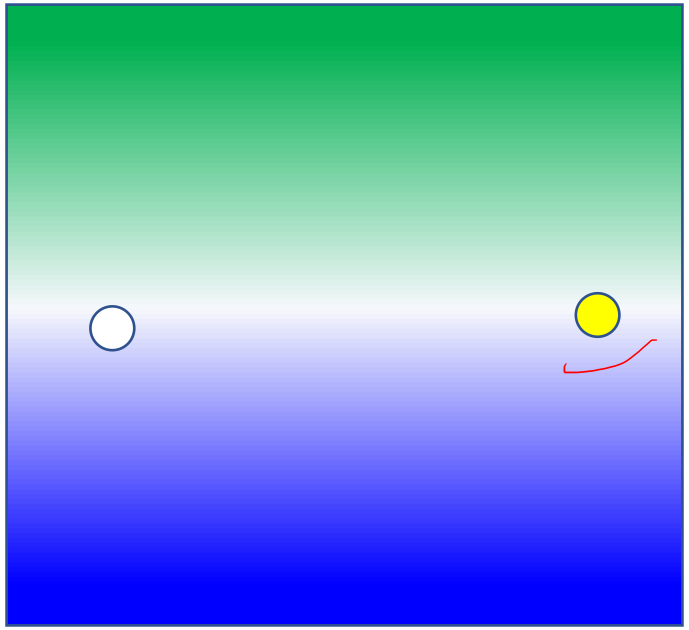
If we choose H, $P(H)$, how can we generate instances of event/evidences $P(E|H)$
If we are in H4: Winter and in Canada, generate a day → it would be mostly no sun!

LR is Discriminative!

Logistic Regression

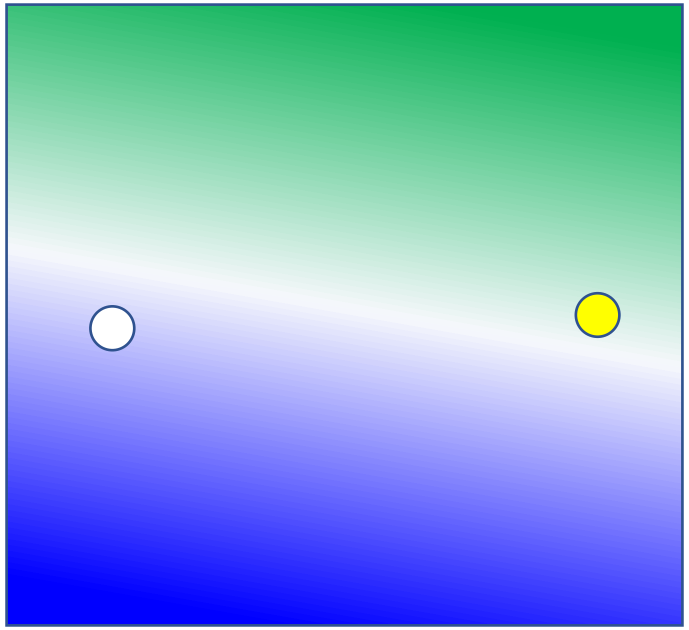
Don’t care about hypothesis and their probabilities.
Try to maximize the distance of data from different classes.
Try to discriminate the most!

Logistic Regression



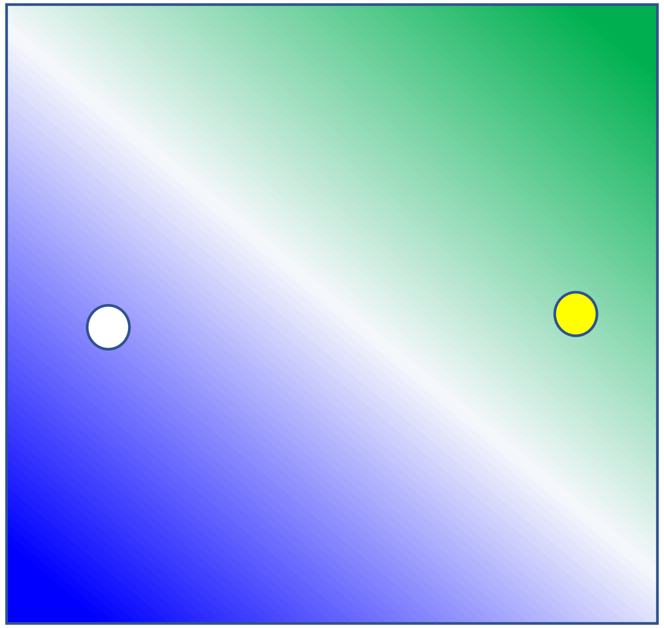
$$P(+|x_+)=0.50 \quad P(-|x_+)=0.50$$
$$P(-|x_-)=0.50 \quad P(+|x_-)=0.50$$

Logistic Regression



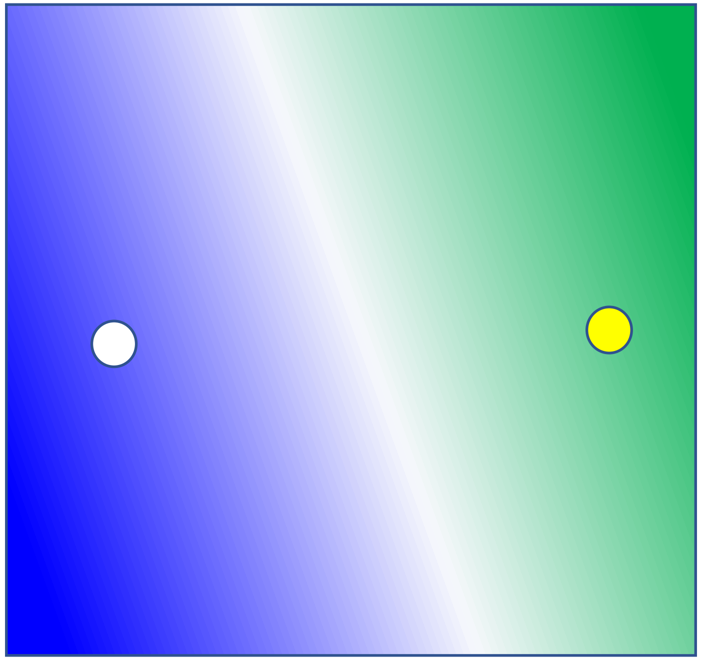
$P(+|x_+)=0.55$ $P(-|x_+)=0.45$
$P(-|x_-)=0.55$ $P(+|x_-)=0.45$

Logistic Regression



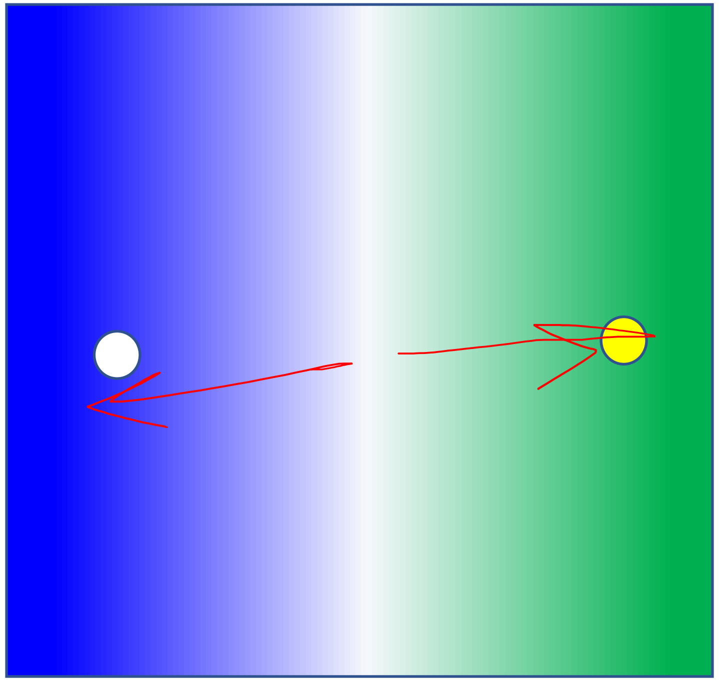
$P(+|x_+)=0.65$ $P(-|x_+)=0.35$
$P(-|x_-)=0.65$ $P(+|x_-)=0.35$

Logistic Regression



$P(+|x_+)=0.75$ $P(-|x_+)=0.25$
$P(-|x_-)=0.75$ $P(+|x_-)=0.25$

Logistic Regression



$P(+|x_+)=0.85$ $P(-|x_+)=0.15$

$P(-|x_-)=0.85$ $P(+|x_-)=0.15$

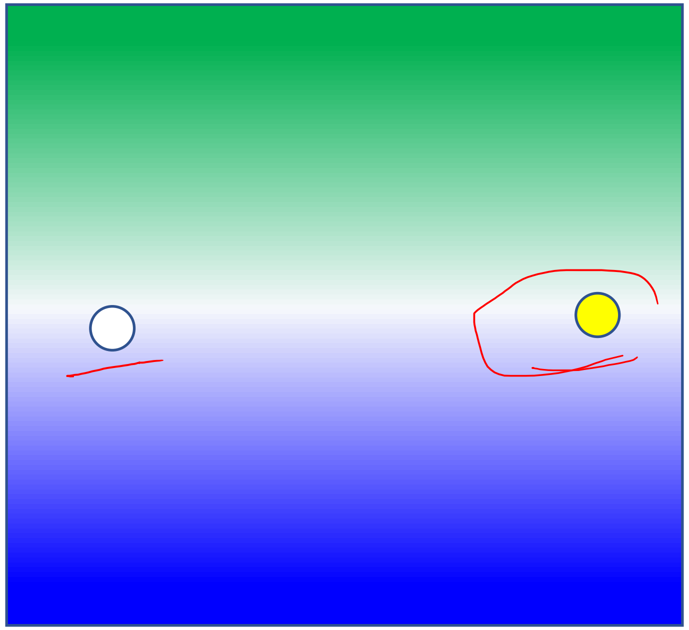
Logistic Regression

(I) Iterative Process (online vs. off)

(II) Optimization = Discriminate classes the most

Logistic Regression

More for $P(-|x_-) \uparrow$



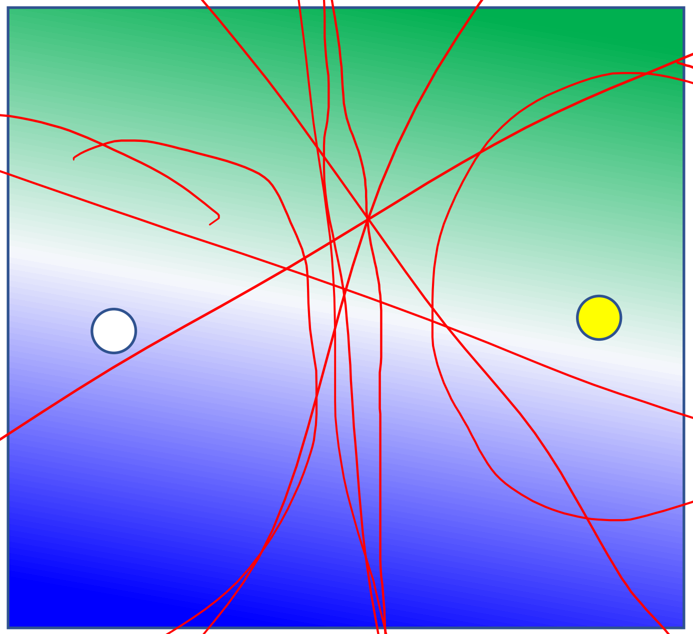
More for $P(+|x_+) \uparrow$

$P(+|x_+)=0.50$ $P(-|x_+)=0.50$

$P(-|x_-)=0.50$ $P(+|x_-)=0.50$

Logistic Regression

Even more for $P(-|x_-) \uparrow$



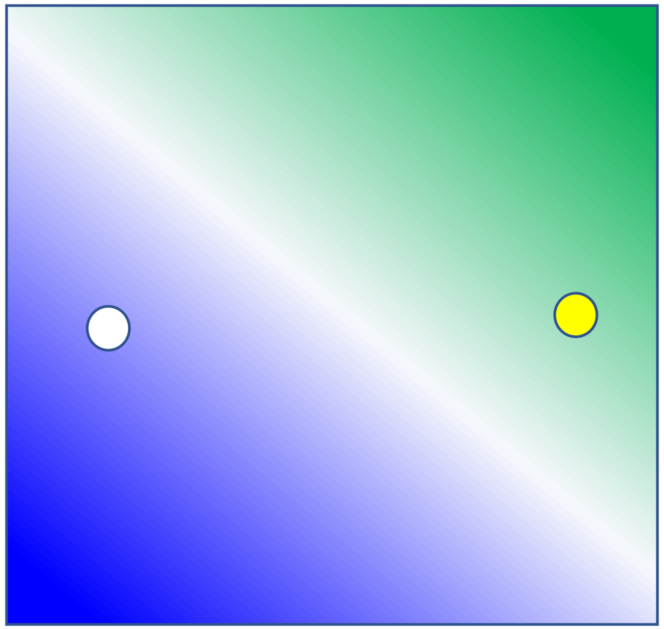
Even more for $P(+|x_+) \uparrow$

$P(+|x_+)=0.55$ $P(-|x_+)=0.45$

$P(-|x_-)=0.55$ $P(+|x_-)=0.45$

Logistic Regression

Even more for $P(-|x_-)\uparrow$



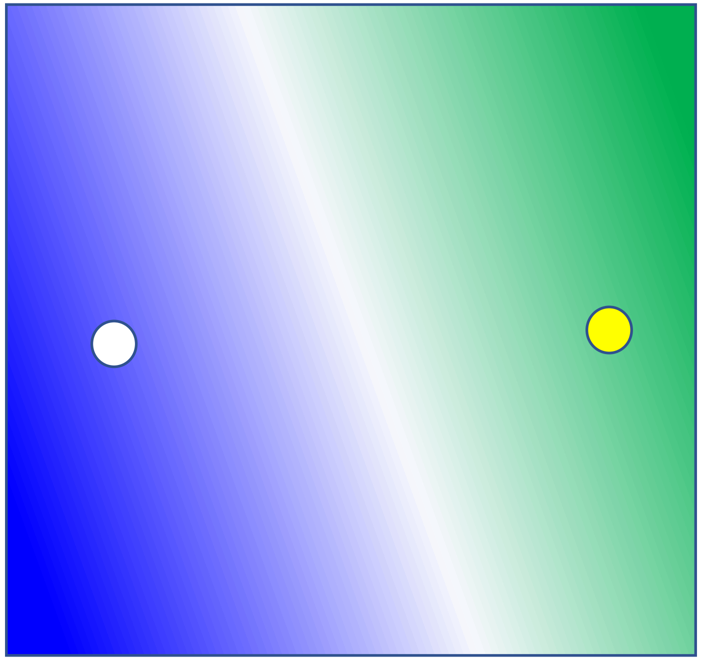
Even more for $P(+|x_+)\uparrow$

$P(+|x_+)=0.65$ $P(-|x_+)=0.35$

$P(-|x_-)=0.65$ $P(+|x_-)=0.35$

Logistic Regression

Even more for $P(-|x_-)\uparrow$



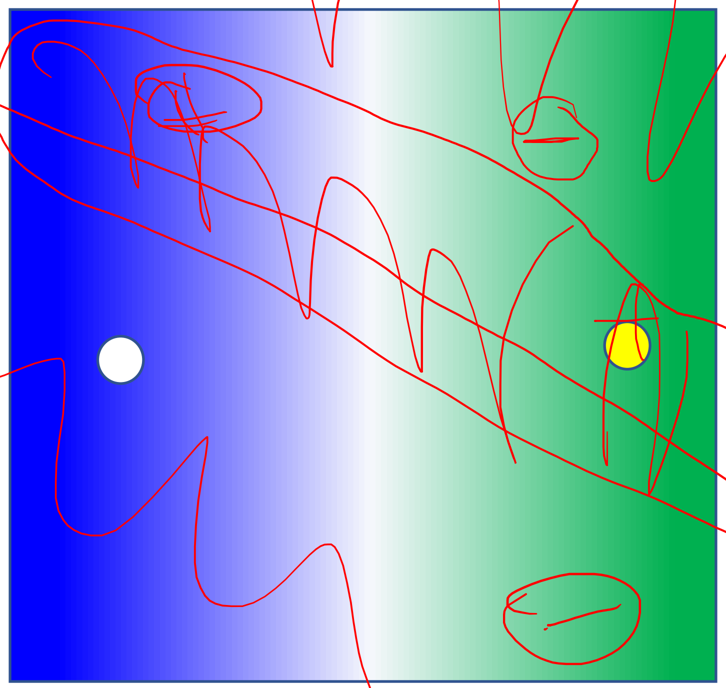
Even more for $P(+|x_+)\uparrow$

$P(+|x_+)=0.75$ $P(-|x_+)=0.25$

$P(-|x_-)=0.75$ $P(+|x_-)=0.25$

Logistic Regression

Max for $P(-|x_-)\uparrow$



Max for $P(+|x_+)\uparrow$

$P(+|x_+)=0.85$ $P(-|x_+)=0.15$

$P(-|x_-)=0.85$ $P(+|x_-)=0.15$

Logistic Regression

(I) Iterative Process: We can change the line that discriminate

(II) Optimization = Discriminate classes the most

(I) Maximizing the $P(+|x_+) + P(-|x_-)$

TP TN

Logistic Regression

(I) What is that line?

(II) What if we have more than two inputs?

$P(A,B,C,D \ldots) = P(A)P(B) \ldots$

$P(AB) = P(A)P(B|A)$

$P(A)P(B|A)P(C|AB) \ldots$

Logistic Regression

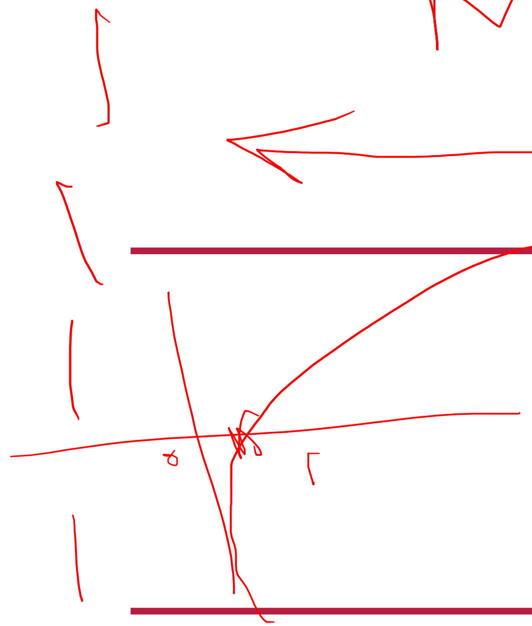
$= P(A)P(B)$ assumption → independence

n train

(I) Iterative Process: We can change the line that discriminate

(II) Optimization

(I) Max ($\prod_{x\in\{+\}} P(+|x_+) \prod_{x\in\{-\}} P(-|x_-)$)



$M \Rightarrow$

$M(x^1) = 0.3 \quad \log(0.3) + \quad -2000$

$M(x^2) \quad \log 0.8 \quad -10$

$+$

$M(x^3) \quad \log 0.5 \quad -1000$

Logistic Regression

$M(x^4) \quad \log 0.2 \quad -3000$

$56 \downarrow 6$

(I) Iterative Process: We can change the line that discriminate

(II) Optimization

(I) $\text{Max}\left(\sum_{x\in\{+\}} \text{Log } P(+|x_+) + \sum_{x\in\{-\}} \text{Log } P(-|x_-)\right)$

Logistic Regression

(I) Iterative Process: We can change the line that discriminate

(II) Optimization

(I) Min $-(\sum_{x \in \{+\}} \text{Log } P(+|x_+) + \sum_{x \in \{-\}} \text{Log } P(-|x_-))$

Logistic Regression

c= {-, +} → y={0, 1} 2, 3 ...

(I) Iterative Process: We can change the line that discriminate

(II) Optimization

(I) Min $-\sum_{(x,y)\in D}[(y)\text{Log } P(y|x_y) + (1-y)\text{ Log } P(y|x_y)]$

1 P(1|X1)

0 × ———

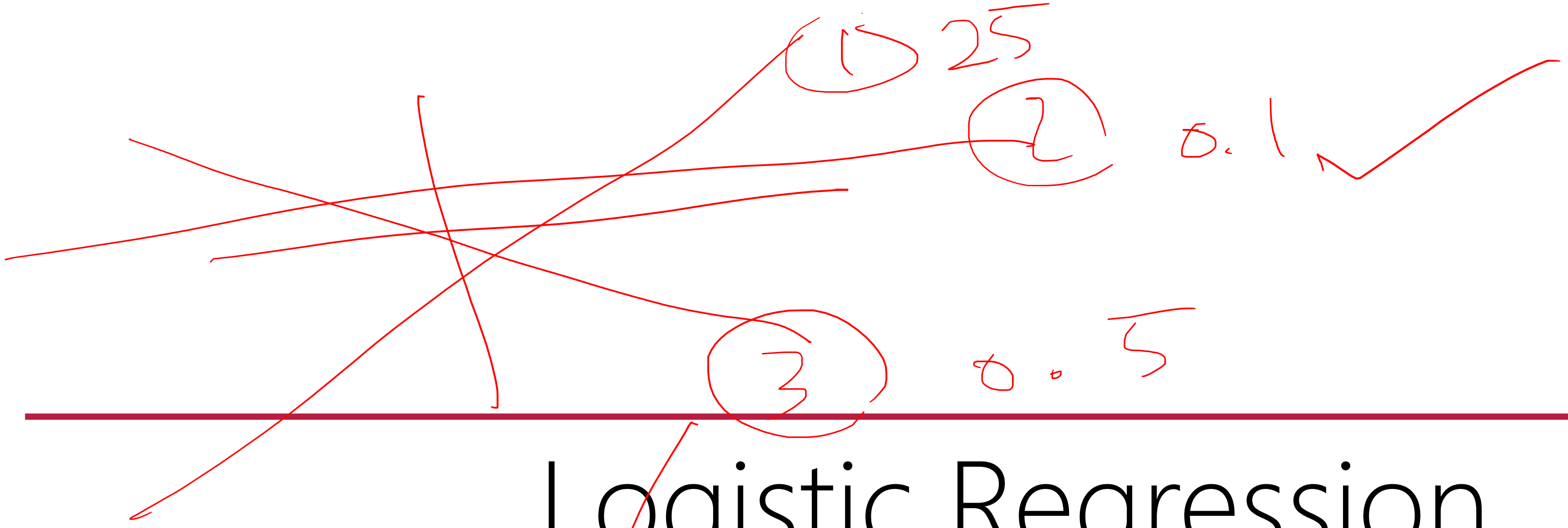
1 × log(0|x0)

Logistic Regression
c= {-, +} → y={0, 1}

(I) Iterative Process: We can change the line that discriminate

(II) Optimization

(I) Min $-\sum_{(x,y)\in D} \text{Log } P(y|x_y)]$



Logistic Regression

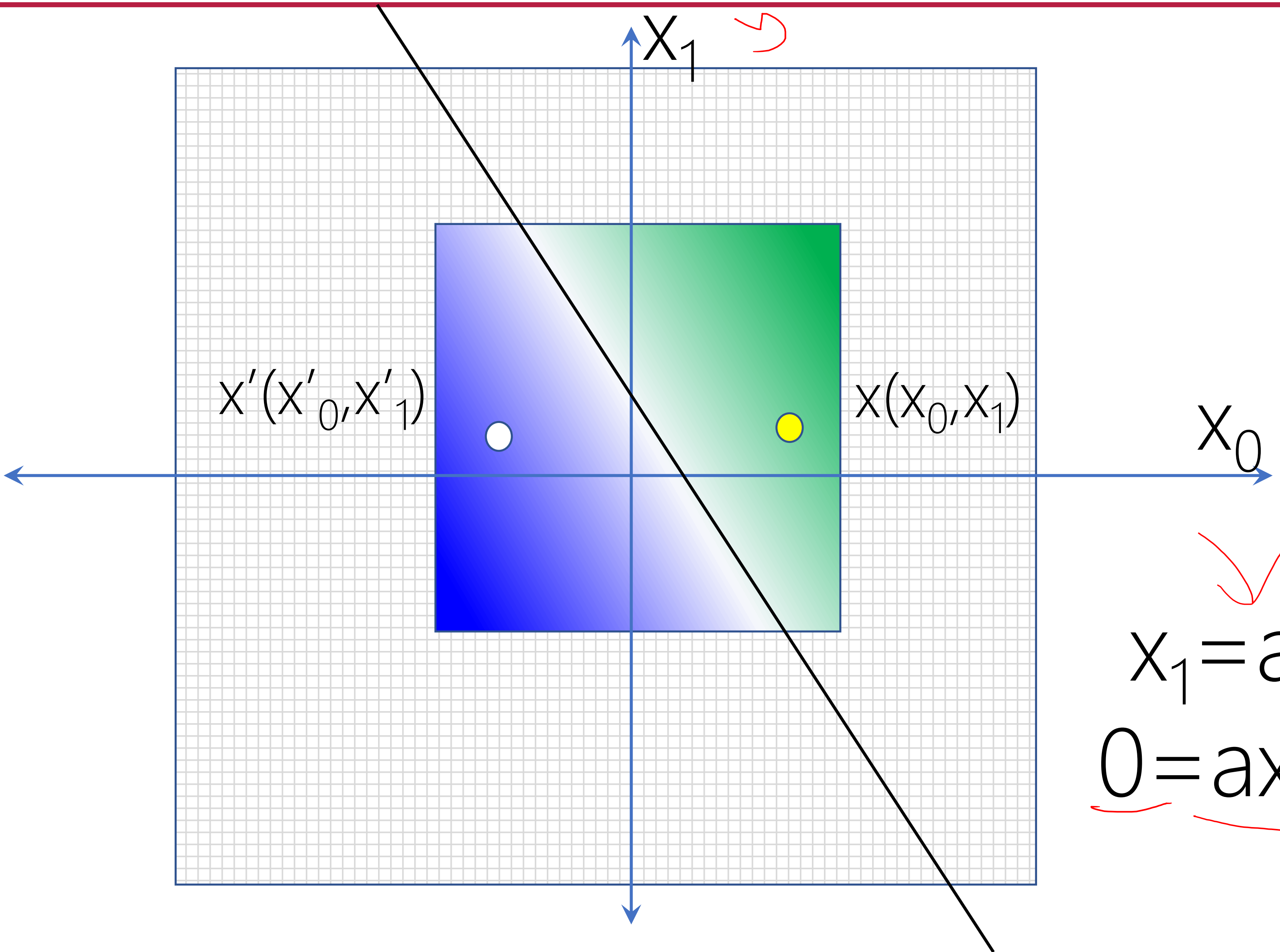
$c = \{-, +\} \rightarrow y = \{0, 1\}$

(I) Iterative Process: We can change the line that discriminate

(II) Optimization

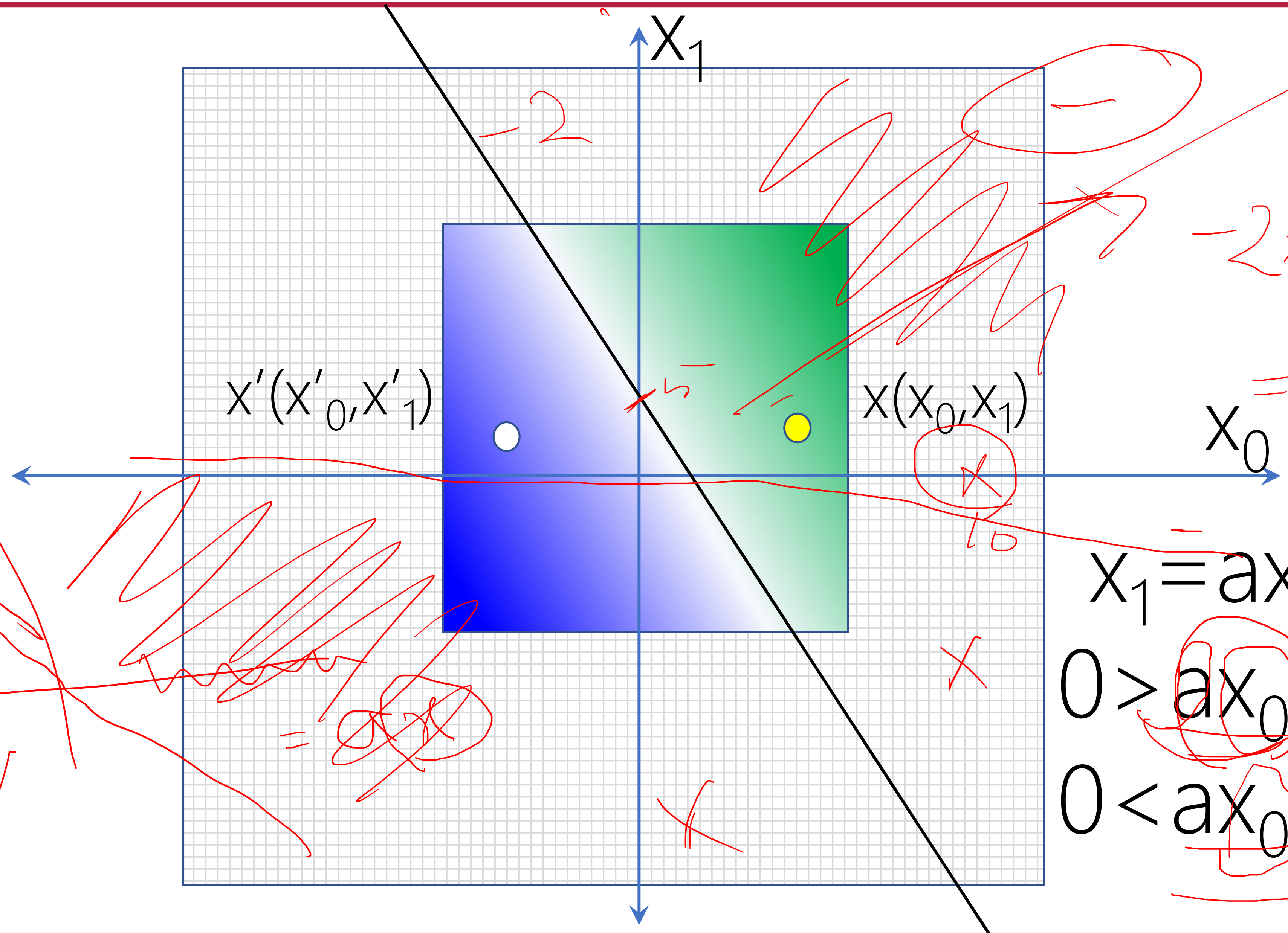
(I) $\text{Min} -\sum_{(x,y)\in D} [(y) \text{Log } P(+|x_+) + (1-y) \text{ Log } P(-|x_-)]$

Logistic Regression



$$x_1 = ax_0 + b$$
$$0 = ax_0 - x_1 - b$$

Logistic Regression

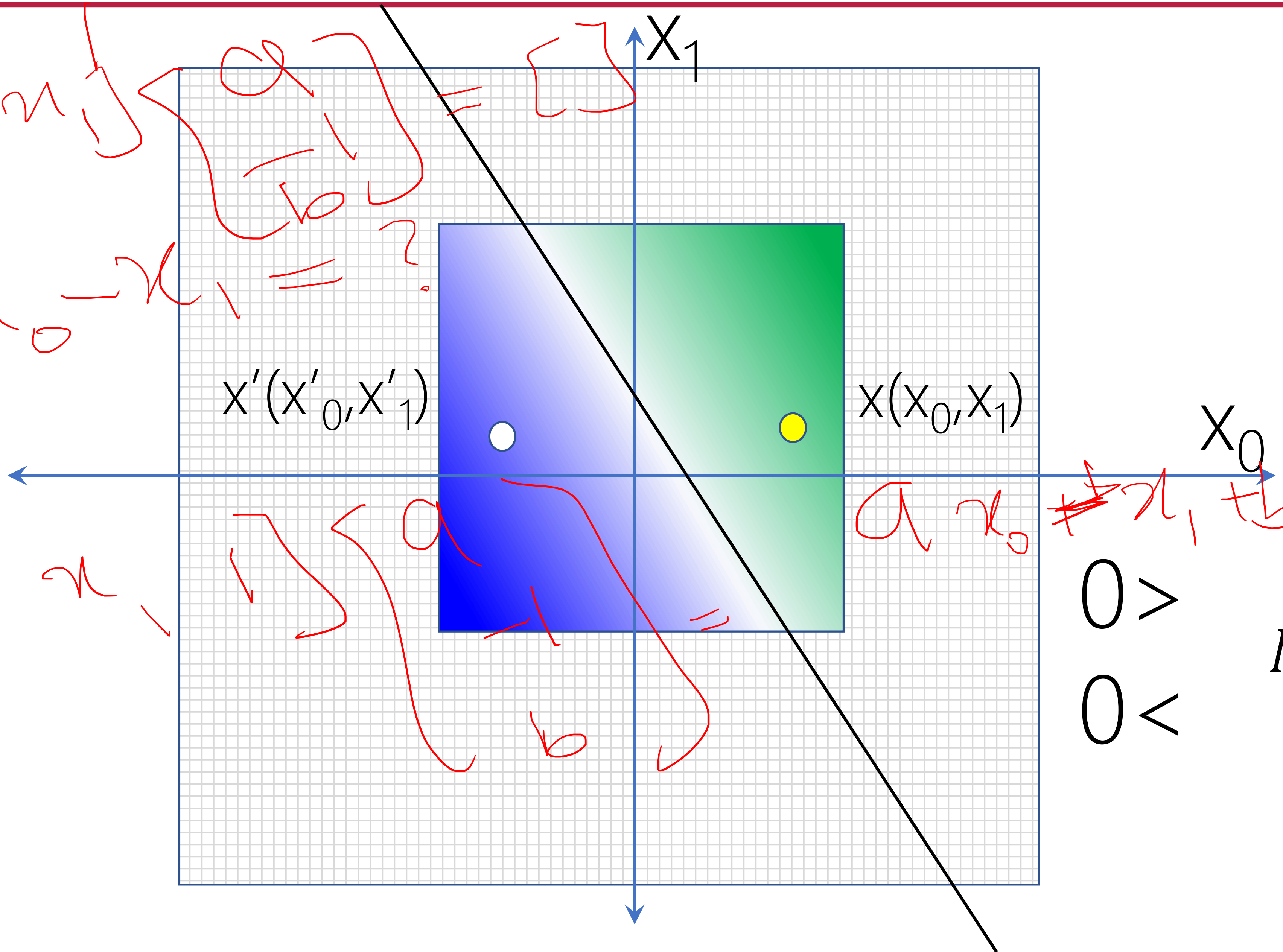


$$x_1 = ax_0 + b$$

$$0 > ax_0 - x_1 + b$$

$$0 < ax_0 - x_1 + b$$

Logistic Regression



$[x_0 \; x_1]\begin{bmatrix} a \\ b \end{bmatrix} = [\;]$

$a x_0 - x_1 = ?$

$[x_0 \; x_1]\begin{bmatrix} a \\ b \end{bmatrix}$

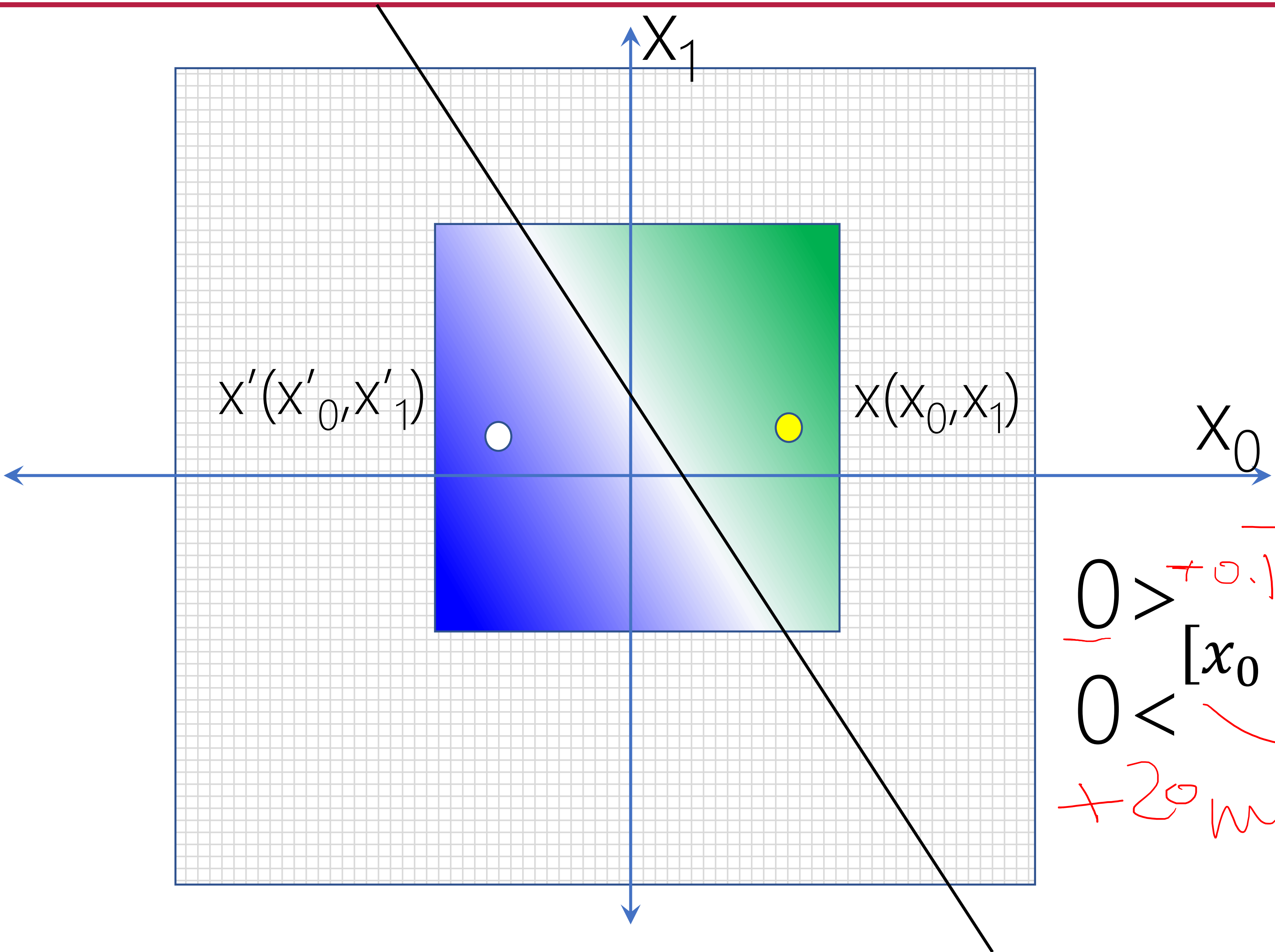
$a x_0 + x_1 + b$

0>

0<

$M_1 M_2$

Logistic Regression



$-\infty$

$+0.1$

$$0 > [x_0\ x_1\ \ 1]\begin{bmatrix} a \\ -1 \\ b \end{bmatrix}$$

$$0 < [x_0\ x_1\ \ 1]\begin{bmatrix} a \\ -1 \\ b \end{bmatrix}$$

$+20m^{0.2}$

Logistic Regression

$$\begin{matrix} 0 > \\ 0 < \end{matrix} \quad \underbrace{[x_0\ x_1}_{x_0\ x_1\ x_2}\ \ldots x_d\ 1]\begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_{d+1} \end{bmatrix} = [?]$$

$x_0\ x_1 \cdots x_n$

(n)

d = 1 → Line in 2-dimension

d = 2 → Plane in 3 dimension

d = n → Hyperplane in (n+1) dimension

Parametric vs. non-Parametric

LR vs. Naïve Bayes

(True Bayesian Inference is Parametric, Why?)

Logistic Regression

$$0 > f(X) \longrightarrow -\infty$$

$$0 < f(X) \longrightarrow +\infty$$

d = 1 → Line in 2-dimension

d = 2 → Plane in 3 dimension

d = n → Hyperplane in (n+1) dimension

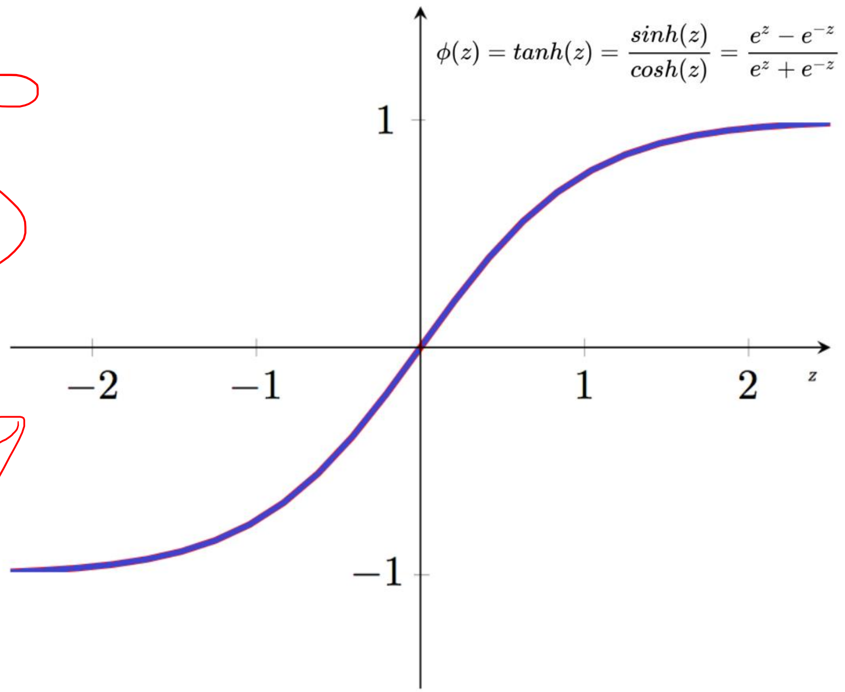
Logistic Regression: Squish by Tanh

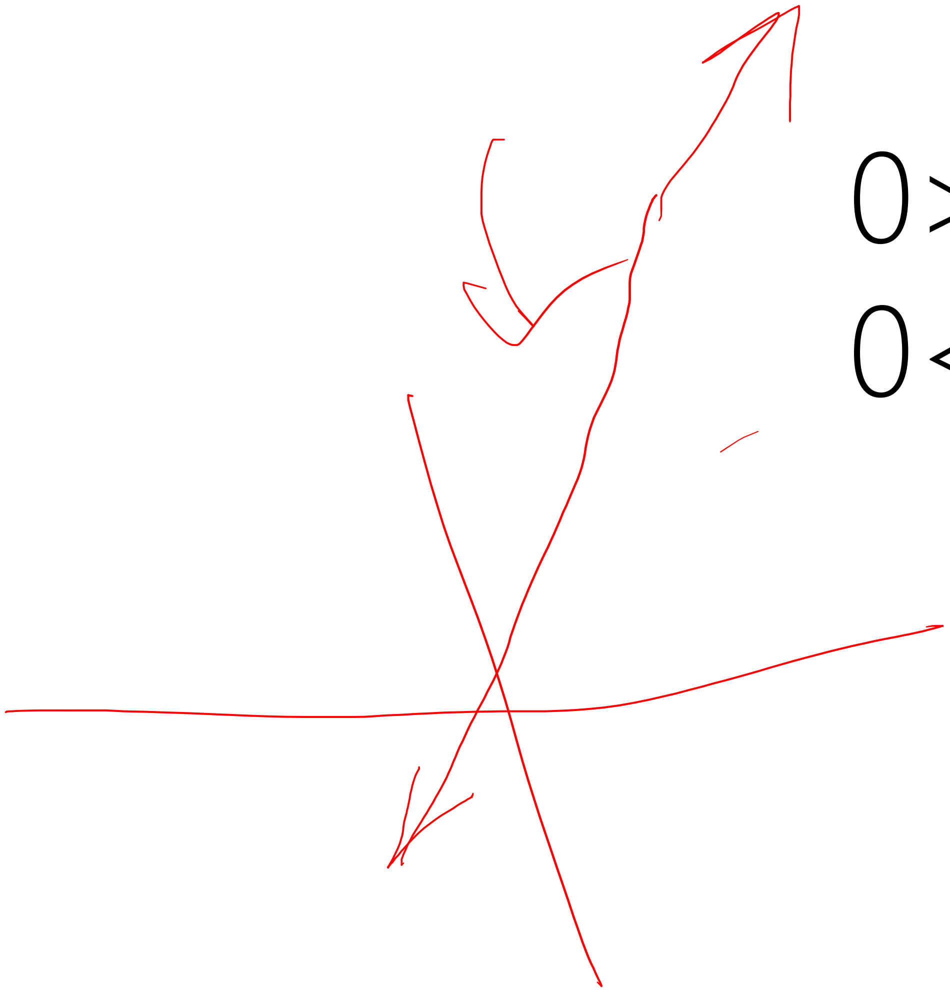
[-1, +1]

$-\infty$

$$0 > \; f(X) \longrightarrow -1$$

$$0 < \; f(X) \longrightarrow +1$$

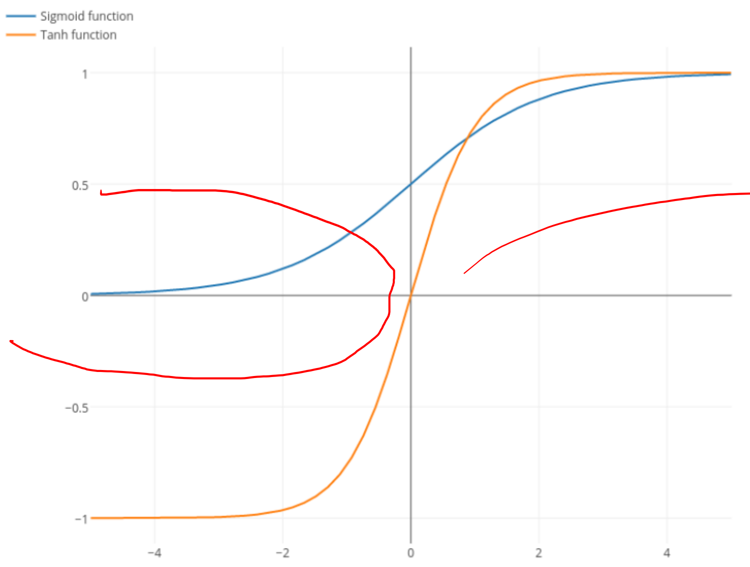




Logistic Regression: Squish by Sigmoid

$$0 > \quad f(X) \longrightarrow 0$$

$$0 < \quad f(X) \longrightarrow +1$$



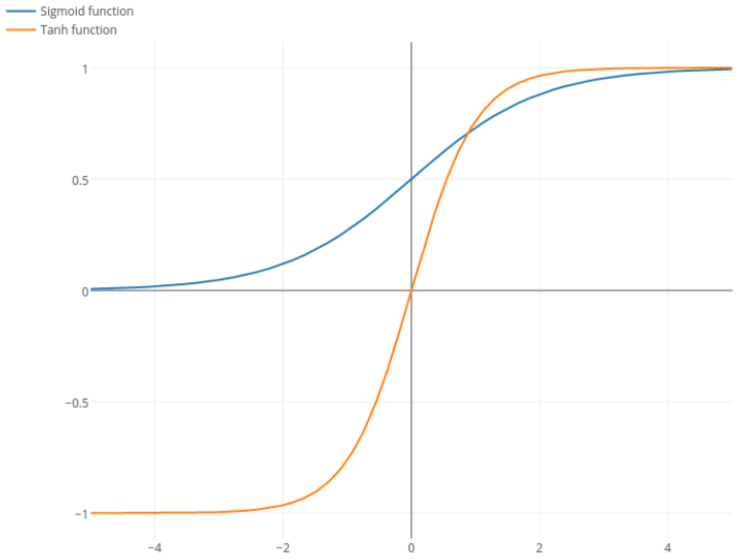
$$S(f(X)) = \frac{1}{1+e^{-f(X)}}$$

Becomes very similar to probability values!

Logistic Regression: Squish

$0 > \quad f(X) \longrightarrow 0$

$0 < \quad f(X) \longrightarrow +1$



$$S(f(X)) = \frac{1}{1+e^{-f(X)}}$$

Becomes very similar to probability values!

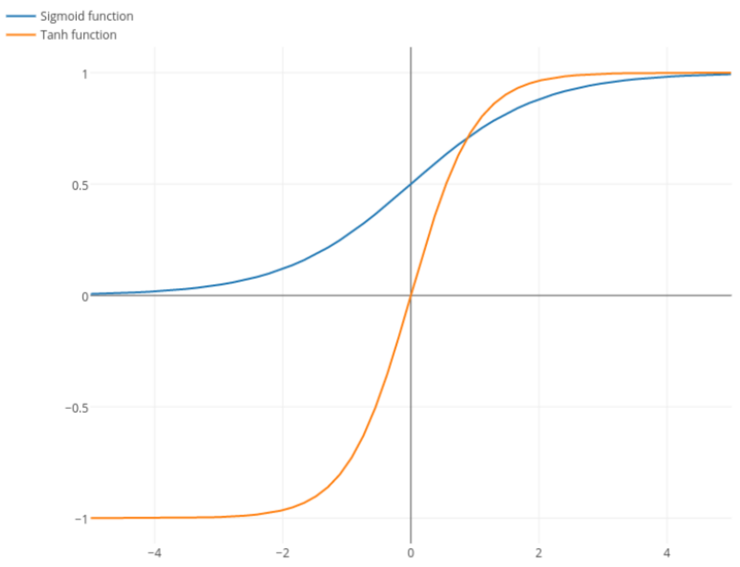
But only for positive class (+)

y=1 → P(y|x) = P(+|x) = Sigmoid(x)

Logistic Regression: Squish

$$0 > \quad f(X) \longrightarrow \mathbf{0}$$

$$0 < \quad f(X) \longrightarrow +\mathbf{1}$$



$$S(f(X)) = \frac{1}{1+e^{-f(X)}}$$

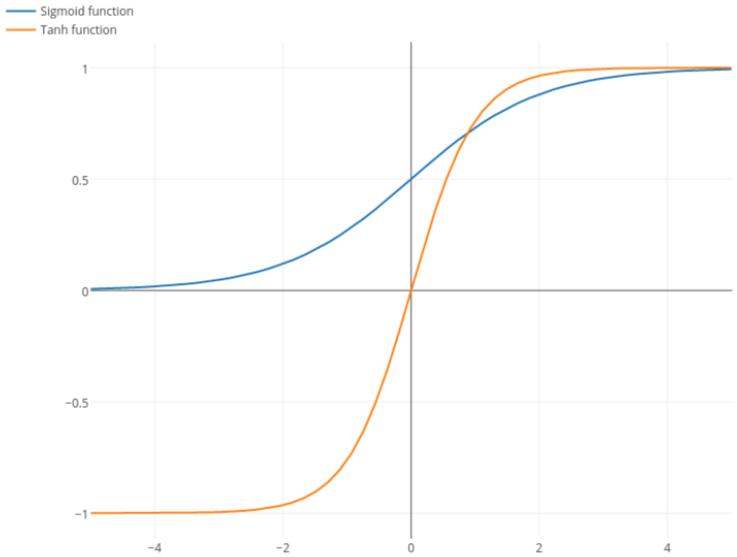
Becomes very similar to probability values!

For negative class (-)?

Logistic Regression: Squish

$$0 > \quad f(X) \longrightarrow \mathbf{0}$$

$$0 < \quad f(X) \longrightarrow \mathbf{+1}$$



$$S(f(X)) = \frac{1}{1+e^{-f(X)}}$$

Becomes very similar to probability values!

For negative class (-)

$P(+|x) + P(-|x) = 1 \rightarrow P(-|x) = 1 - P(+|x)$

$x_0 \quad x_1$

$\begin{bmatrix} 2 \\ -1 \\ 5 \\ 10 \\ 1 \\ 50 \end{bmatrix}$

$1 - \text{Sigmoid}$

$-[\text{Sigmoid}]$

Logistic Regression

c= {-, +} → y={0, 1}

(I) Iterative Process: We can change the *f* that discriminate

x

(II) Optimization

(I) Min $-\sum_{(x,y)\in D} \text{Log } P(y|x_y)] \rightarrow \begin{cases} y = 1: P(+|x_+) = Sigmoid(f(x_+)) \\ y = 0: P(-|x_-) = 1 - Sigmoid(f(x_+)) \end{cases}$

Logistic Regression
P ~ f

(I) Iterative Process: We can change the f that discriminate

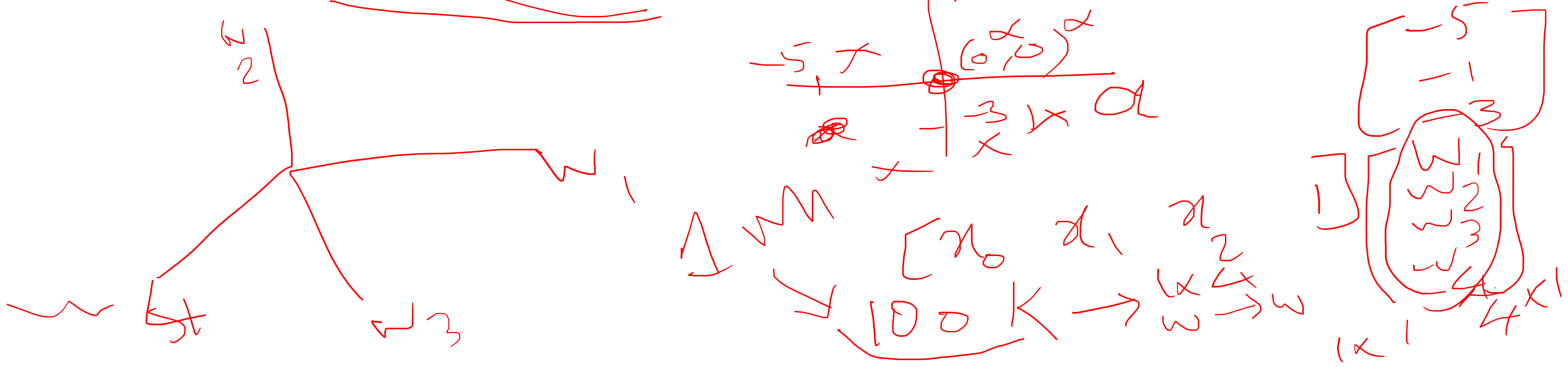
(II) Optimization

(I) Min $-\sum_{(x,y)\in D} \text{Log } P(y|x_y)]$ → $\begin{cases} y = 1: P(+|x_+) = Sigmoid(f(x_+)) \\ y = 0: P(-|x_-) = 1 - Sigmoid(f(x_+)) \end{cases}$

Logistic Regression

Optimization: Min $-\sum_{(x,y)\in \mathbf{D}} \text{Log } P(y|x_y)]$ $\rightarrow$ $\begin{cases} y = 1:\ P(+|x_+) = Sigmoid(f(x_+)) \\ y = 0: P(-|x_-) = 1\ -\ Sigmoid(f(x_+)) \end{cases}$

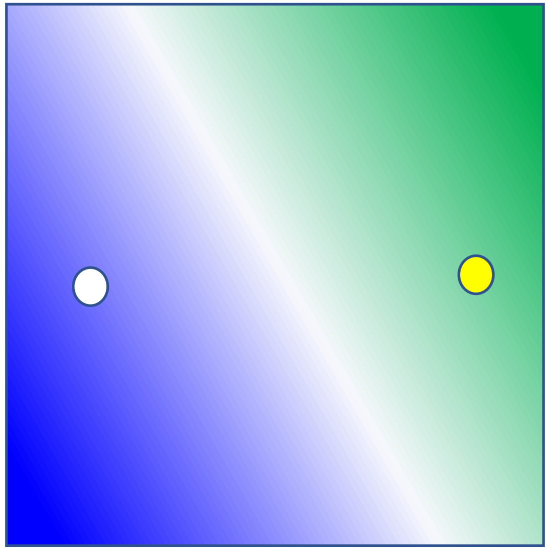
- Derivatives per function weights (Gradients)
  - Update on each input data
  - Update on batches of input data (Stochastic Gradient Descend)
  - Update on multiple rounds (epoch) of ALL data



Logistic Regression

Optimization: Min $-\sum_{(x,y)\in D}$ Log P(y|x_y)] → $\begin{cases} y = 1: \; P(+|x_+) = Sigmoid(f(x_+)) \\ y = 0: P(-|x_-) = 1 - Sigmoid(f(x_+)) \end{cases}$

- Function *f* is linear function of weights

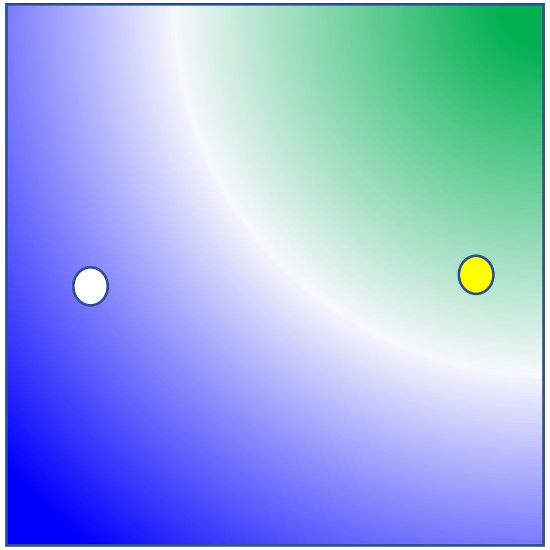


$$[x_0 \; x_1 \; \ldots x_d \, 1]\begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_{d+1} \end{bmatrix} = [?]$$

Logistic Regression

Optimization: Min $-\sum_{(x,y)\in D}$ Log P(y|x_y)] → $\begin{cases} y = 1: \; P(+|x_+) = Sigmoid(f(x_+)) \\ y = 0: P(-|x_-) = 1 - Sigmoid(f(x_+)) \end{cases}$

- Can we have *f* as a non-linear function of weights?

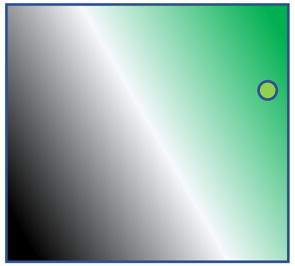


Multiclass with Logistic Regression

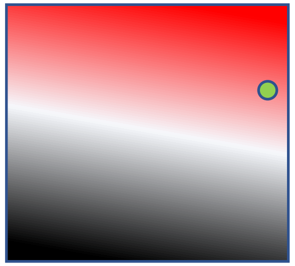
Multi-target (Multi-label)

Optimization: Min $-\sum_{(x,y)\in D} \text{Log}\ P(y|x_y)]$

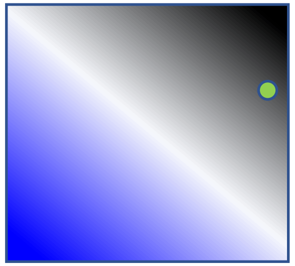
- Data point can be member of more than one class



green class vs. else: $[x_0\ x_1\ \ldots x_d\ 1]\begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_{d+1} \end{bmatrix} = [?]$



red class vs. else: $[x_0\ x_1\ \ldots x_d\ 1]\begin{bmatrix} w'_1 \\ w'_2 \\ \ldots \\ w'_{d+1} \end{bmatrix} = [?]$



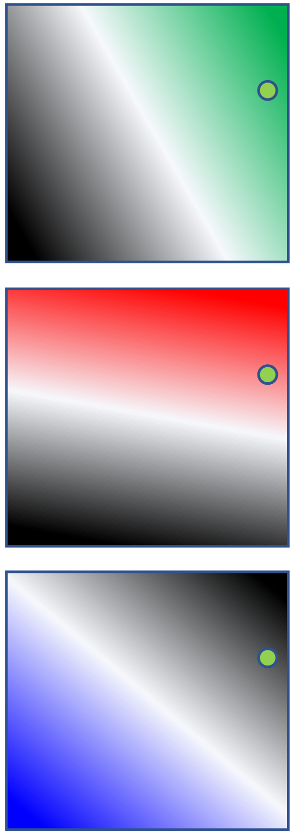
blue class vs. else: $[x_0\ x_1\ \ldots x_d\ 1]\begin{bmatrix} w''_1 \\ w''_2 \\ \ldots \\ w''_{d+1} \end{bmatrix} = [?]$

Multiple Binary Classification

Multi-target (Multi-label)

Optimization: Min $-\sum_{(x,y)\in D} \text{Log } P(y|x_y)]$

- Data point can be member of more than one class



green

not red

not blue

$$[x_0\ x_1\ \dots x_d\ 1]\begin{bmatrix} w_1 & w'_1 & w''_1 \\ w_2 & w'_2 & w''_2 \\ \dots & \dots & \dots \\ w_{d+1} & w'_{d+1} & w''_{d+1} \end{bmatrix} = [?\,,?\,,?\,] = [1,0,1]$$

Si

([0,1], [0,1], [0,1])

Multinomial Logistic Regression

Logistic Regression → Softmax Regression

Softmax Regression

$v \longrightarrow e^{v} \longrightarrow \frac{4e^{v}}{|e^{v}|}$

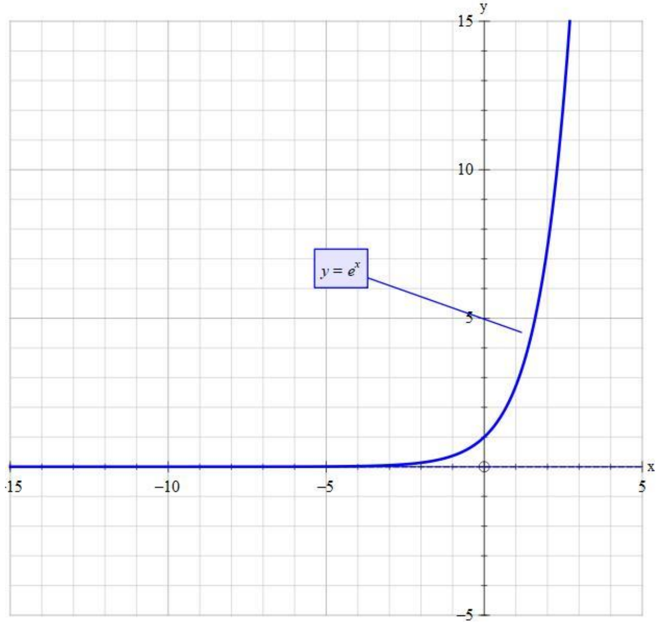
$$\underline{softmax}(\boldsymbol{z} = \underline{[z_1, z_2, \dots, z_d]}) = \frac{1}{\underline{\sum_{i=1}^{d} e^{z_i}}} [e^{z_1}, e^{z_2}, \dots, e^{z_d}]$$

$$0 \le \frac{e^{z_j}}{\sum_{i=1}^{d} e^{z_i}} \le 1$$

$$\sum_{j=1}^{d} \frac{e^{z_j}}{\sum_{i=1}^{d} e^{z_i}} = 1$$

acts like

```
In[2]: import torch
In[3]: sample_vector = torch.rand(*(1,10))
In[4]: sample_vector
Out[4]:
tensor([[0.6084, 0.8744, 0.8193, 0.8784, 0.2906, 0.7169, 0.0126, 0.8783, 0.3822,
         0.3883]])
In[5]: torch.softmax(sample_vector, dim=1)
Out[5]:
tensor([[0.0985, 0.1285, 0.1216, 0.1290, 0.0717, 0.1098, 0.0543, 0.1290, 0.0786,
         0.0790]])
In[6]: torch.sum(torch.softmax(sample_vector, dim=1))
Out[6]: tensor(1.0000)
```



Multi-target (Multi-label)

Optimization: Min $-\sum_{(x,y)\in D} \mathrm{Log}\ P(y|x_y)]$

$\infty, +\infty$

$[0, 0, \ldots]$

$[0, 1], [0, 1], \ldots$

$$[x_0\ x_1\ \ldots x_d\ 1]\begin{bmatrix} w_1 & w'_1 & w''_1 \\ w_2 & w'_2 & w''_2 \\ \ldots & \ldots & \ldots \\ w_{d+1} & w'_{d+1} & w''_{d+1} \end{bmatrix} = [1.5, -1.4, 10] \Rightarrow Softmax$$

$$\Rightarrow [2.0342e-04, 1.1193e-05, 9.9979e-01]$$

```
In[12]: torch.softmax(torch.as_tensor([1.5, -1.4, 10.0]).view(-1), dim=0)
Out[12]: tensor([2.0342e-04, 1.1193e-05, 9.9979e-01])
```

$$[x_0\ x_1\ \ldots x_d\ 1]\begin{bmatrix} w_1 & w'_1 & w''_1 \\ w_2 & w'_2 & w''_2 \\ \ldots & \ldots & \ldots \\ w_{d+1} & w'_{d+1} & w''_{d+1} \end{bmatrix} = [1.5, -1.4, 10]$$

$$\Rightarrow sigmoid = [0.8176, 0.1978, 1.0000]$$

$$\Rightarrow softmax \Rightarrow [0.3652, 0.1965, 0.4383]$$

```
In[14]: torch.sigmoid(torch.as_tensor([1.5, -1.4, 10.0]).view(-1))
Out[14]: tensor([0.8176, 0.1978, 1.0000])
In[15]: torch.softmax(torch.sigmoid(torch.as_tensor([1.5, -1.4, 10.0]).view(-1)), dim=0)
Out[15]: tensor([0.3652, 0.1965, 0.4383])
```

Multi-target (Multi-label)

Optimization: Min $-\sum_{(x,y)\in \mathbf{D}}$ Log P(y|x_y)]

Multitarget:$(x,[1,0,1]) \overset{?}{\leftrightarrow} (x,[0.8176, 0.1978, 1.0])$

$$(x,\{c1\}) \overset{?}{\leftrightarrow} (x,[0.8176, \blacksquare, \blacksquare])$$

$$(x,\cancel{\{c2\}}) \overset{?}{\leftrightarrow} (x,[\blacksquare, 0.1978, \blacksquare]) \overset{?}{\leftrightarrow} (x,[\blacksquare, 1-0.1978, \blacksquare])$$

$$(x,\{c3\}) \overset{?}{\leftrightarrow} (x,[\blacksquare, \blacksquare, 1.0])$$

– [Log P(c1|x_{c1}) + Log P(c2|x_{c2}) + Log P(c3|x_{c3})]

– [Log (0.81) + Log (1-0.19) + Log (1.0)]

Multiclass

Optimization: Min $-\sum_{(x,y)\in D} \text{Log } P(y|x_y)]$

Multiclass: $(x,[0,1,0]) \overset{?}{\leftrightarrow} (x,[0.3652, 0.1965, 0.4383])$

$(x, \cancel{\{1\}}) \overset{?}{\leftrightarrow} (x,[0.3652, ■, ■]) \overset{?}{\leftrightarrow} (x,[1-0.3652, ■, ■])$

$(x,\{2\}) \overset{?}{\leftrightarrow} (x,[■, 0.1965, ■])$

$(x,\{3\}) \overset{?}{\leftrightarrow} (x,[■, ■, 0.4383]) \overset{?}{\leftrightarrow} (x,[■, ■, 1-0.4383])$

Multiclass

Optimization: Min $-\sum_{(x,y)\in \mathbf{D}} \text{Log } P(y|x_y)]$

Multiclass: $(x,[0,1,0]) \overset{?}{\leftrightarrow} (x,[0.3652,0.1965,0.4383])$

$$\cancel{(x,\{1\}) \overset{?}{\leftrightarrow} (x,[0.3652,\blacksquare,\blacksquare]) \overset{?}{\leftrightarrow} (x,[1-0.3652,\blacksquare,\blacksquare])}$$

$$(x,\{2\}) \overset{?}{\leftrightarrow} (x,[\blacksquare,0.1965,\blacksquare])$$

$$\cancel{(x,\{3\}) \overset{?}{\leftrightarrow} (x,[\blacksquare,\blacksquare,0.4383]) \overset{?}{\leftrightarrow} (x,[\blacksquare,\blacksquare,1-0.4383])}$$

If x belongs to one class, it does not belong to other classes

Also, in softmax, if we increase one element, it reduces other (sum=1)

Multiclass

Optimization: Min $-\sum_{(x,y)\in D}$ Log P(y|x$_y$)]

Multiclass: $(x,[0,1,0]) \overset{?}{\leftrightarrow} (x,[0.3652, 0.1965, 0.4383])$

$$-\left([0,1,0]\log\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = +\infty$$

$$-\left([0,1,0]\log\begin{bmatrix}0.3652\\0.1965\\0.4383\end{bmatrix}\right) =?$$

$$-\left([0,1,0]\log\begin{bmatrix}0\\1\\0\end{bmatrix}\right) = 0$$

Evaluation

Curves

Threshold-based Model

$P(+|\,x) = 1 - P(-|\,x)$

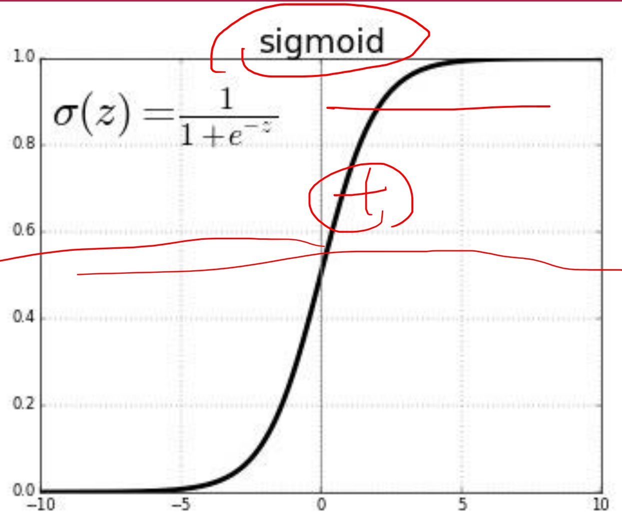
$P(x) = \text{Sigmoid}(f(x))$

$P(x) \geq \delta \longrightarrow$ x is positive

$P(x) < \delta \longrightarrow$ x is negative

~~0.6~~

0.9



0.5

Threshold-based Model
$\delta = 0.5$

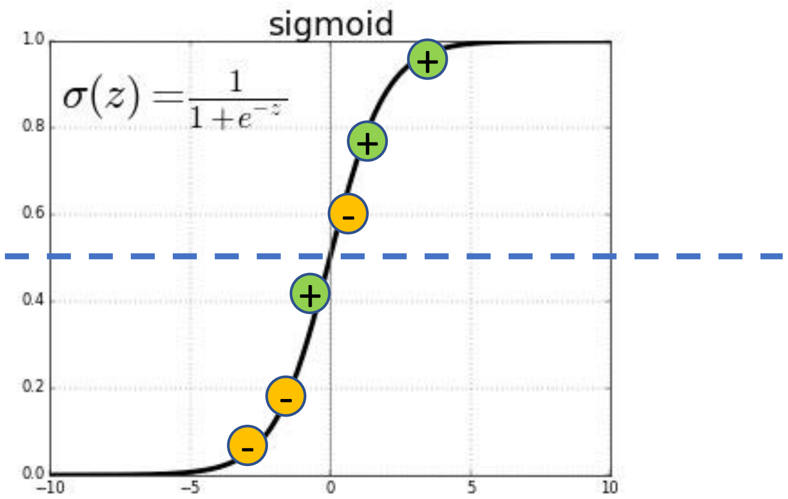
$P(+|\ x) = 1 - P(-|\ x)$

$P(x) = \text{Sigmoid}(f(x))$

$P(x) \geq 0.5 \longrightarrow$ x is positive

$P(x) < 0.5 \longrightarrow$ x is negative

$1 - (\ \ )$



Threshold-based Model

δ=0.0 ⟶ Biased Model ⟶ All are positives

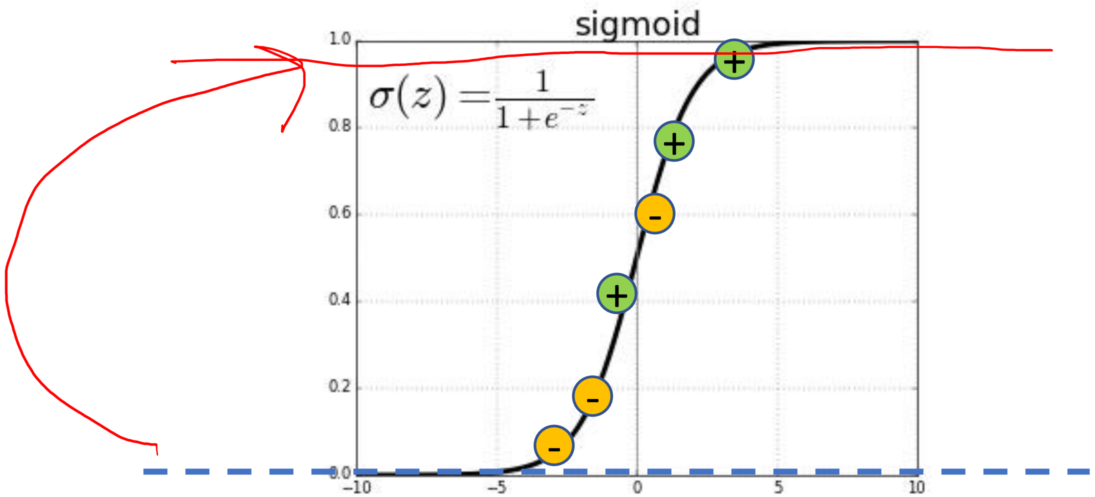
negative

P(+| x) = 1 - P(-| x)

P(x) = Sigmoid(*f*(x))

P(x) ≥ 0.0 ⟶ x is positive

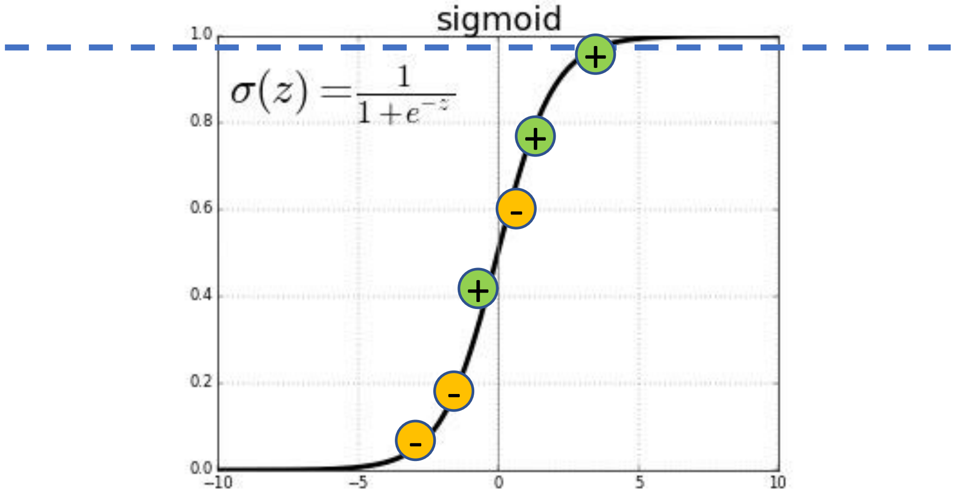
P(x) < 0.0 ⟶ x is negative



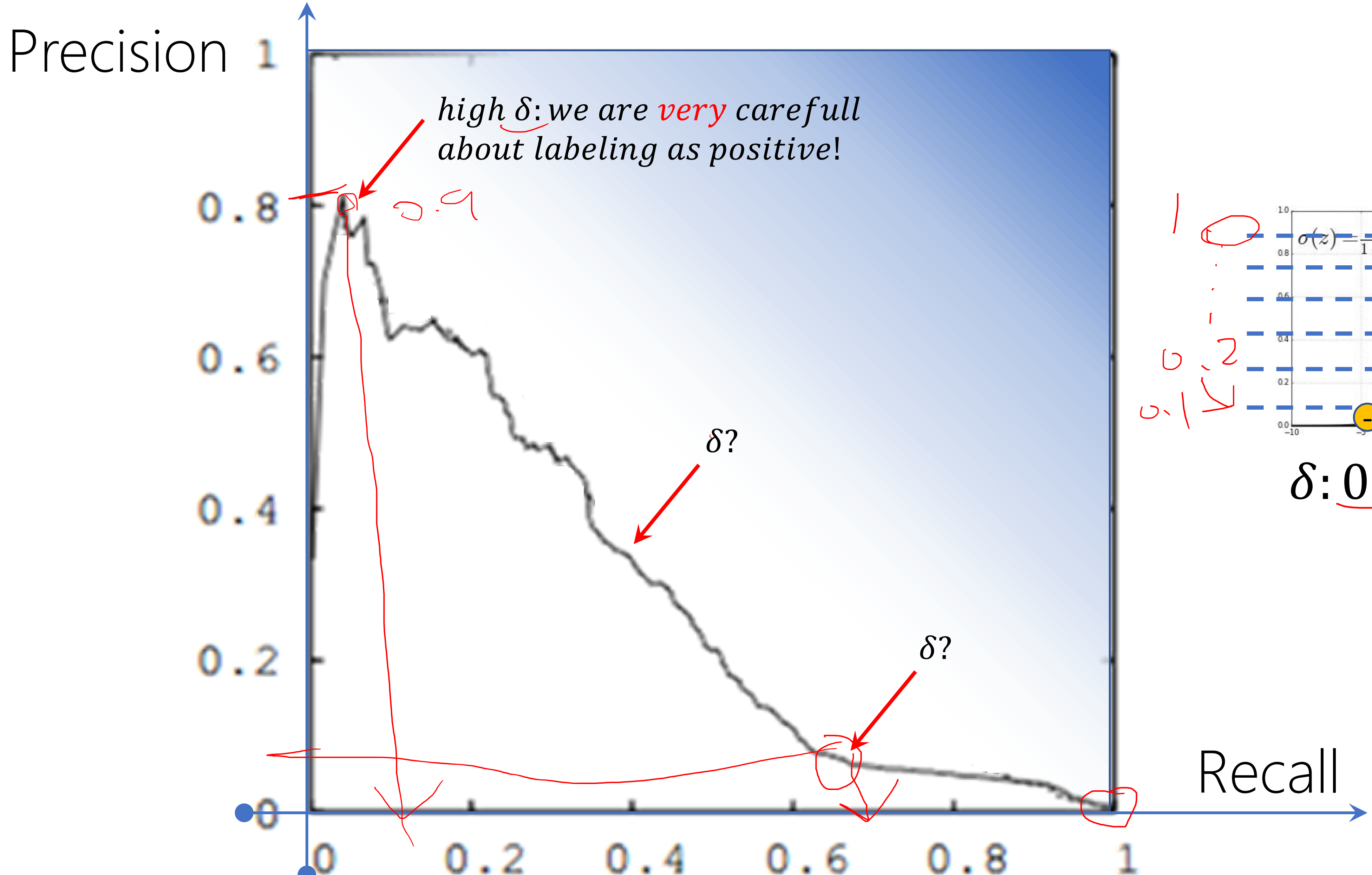
Threshold-based Model

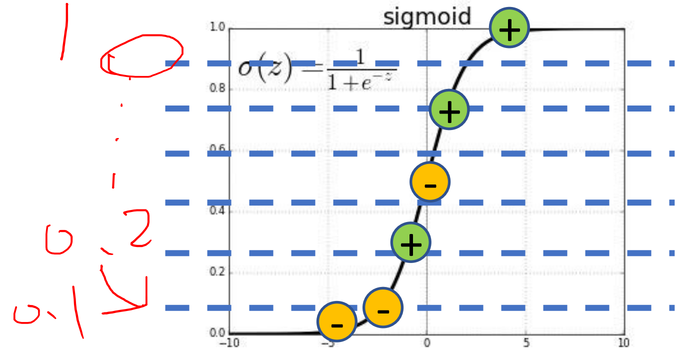
δ=1.0 ⟶ Biased Model ⟶ All are negatives

P(+| x) = 1 - P(-| x)
P(x) = Sigmoid(f(x))
P(x) ≥ 1.0 ⟶ x is positive
P(x) < 1.0 ⟶ x is negative



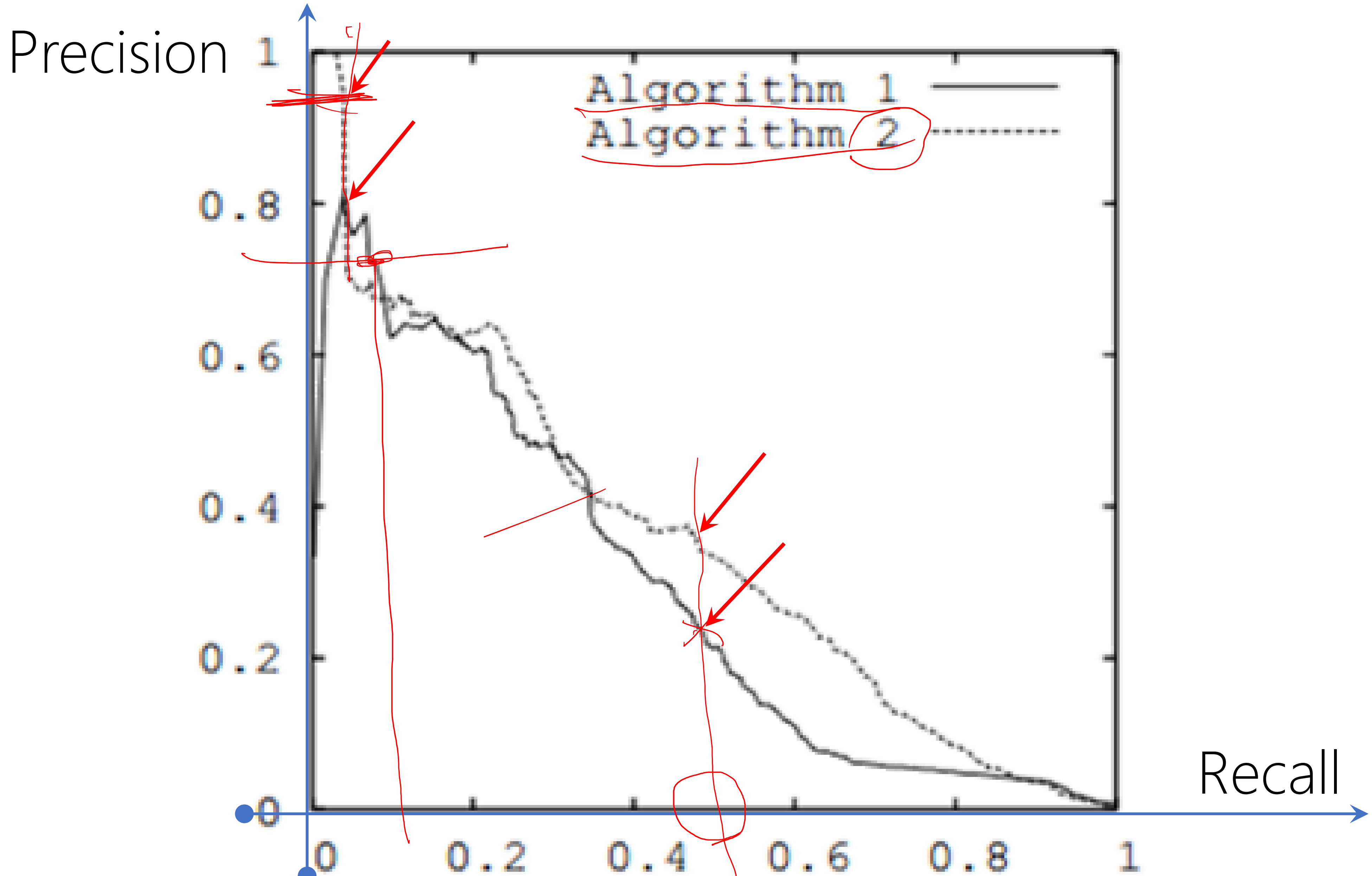
Precision-Recall Curve: Best δ





δ: 0.0 $\rightarrow$ 1.0

Precision-Recall Curve: Model Comparison



Receiver Operating Characteristic ROC

The ROC curve was first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battlefield!

Missile attack vs. passenger airplane!

Recall aka True Positive Rate (TPR)

What percentage of positives are captured.

		gold standard labels		
		gold positive	gold negative	
system output labels	system positive	**true positive**	**false positive**	**precision** $= \frac{tp}{tp+fp}$
	system negative	**false negative**	**true negative**	
		recall $= \frac{tp}{tp+fn}$		**accuracy** $= \frac{tp+tn}{tp+fp+tn+fn}$

Figure 4.4 Contingency table

False Positive Rate (FPR)

What percentage are *incorrectly* captured as positives!

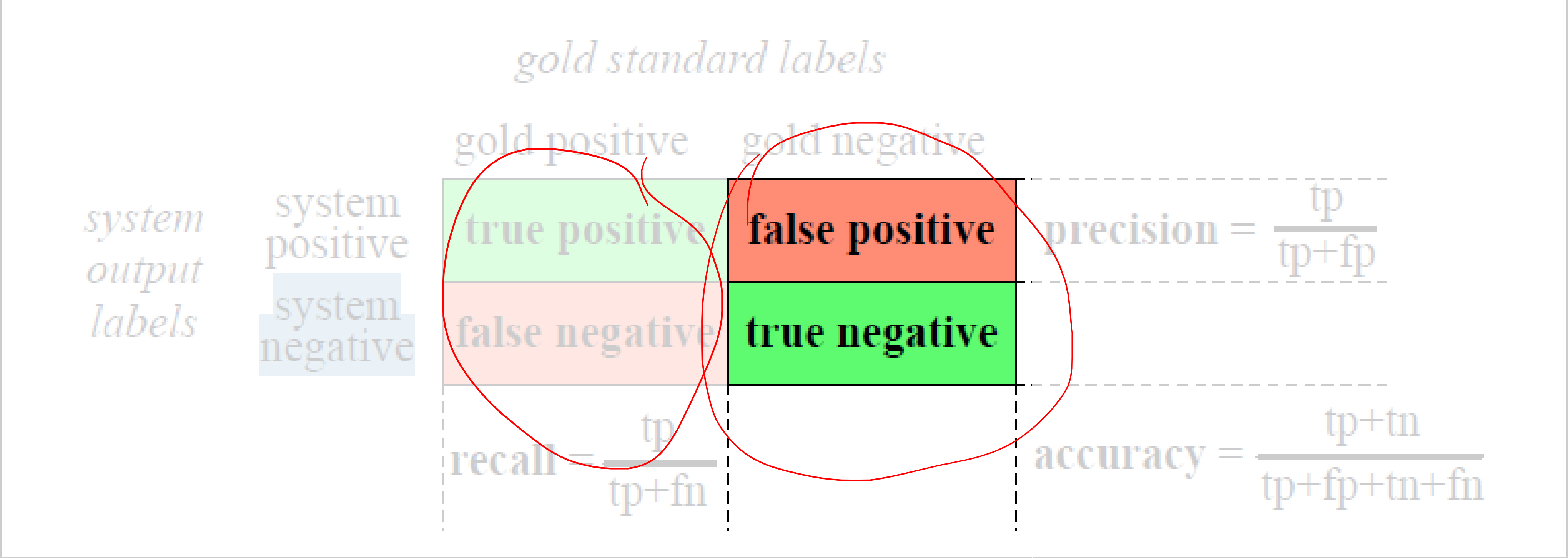


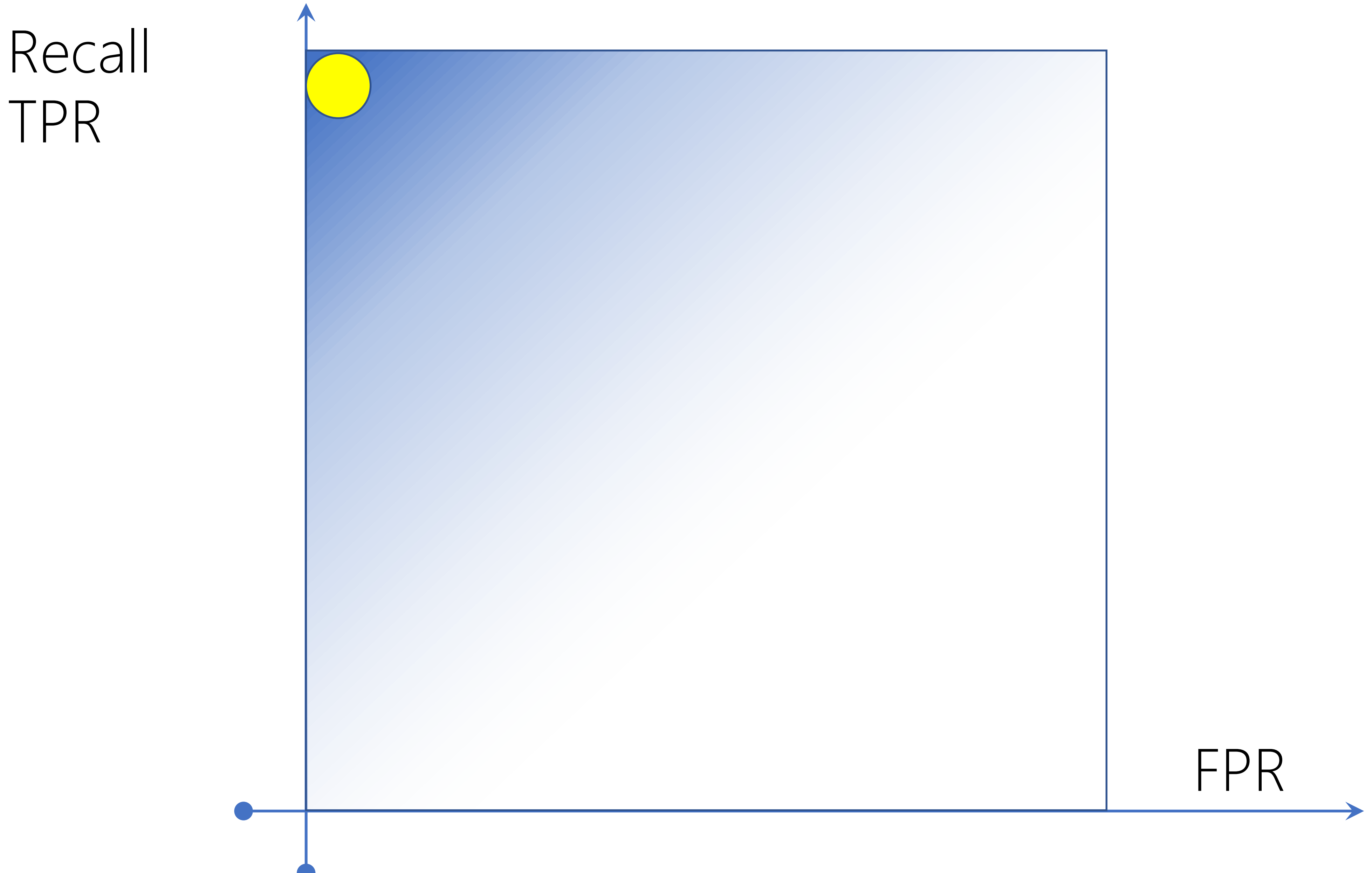
Figure 4.4 Contingency table

Perfect Classifier

	Gold Positive	Gold Negative
Model Positive	N+	0
Model Negative	0	N-

TPR= $\frac{\boldsymbol{N+}}{\boldsymbol{(N+)+0}}$ =1.0 FPR= $\frac{\mathbf{0}}{\boldsymbol{0+(N-)}}$ =0.0

TPR-FPR Curve



Perfect Classifier

	Gold Positive	Gold Negative
Model Positive	N+	0
Model Negative	0	N-

TPR=$\frac{N+}{(N+)+0}$ =1.0 FPR=$\frac{0}{0+(N-)}$ =0.0

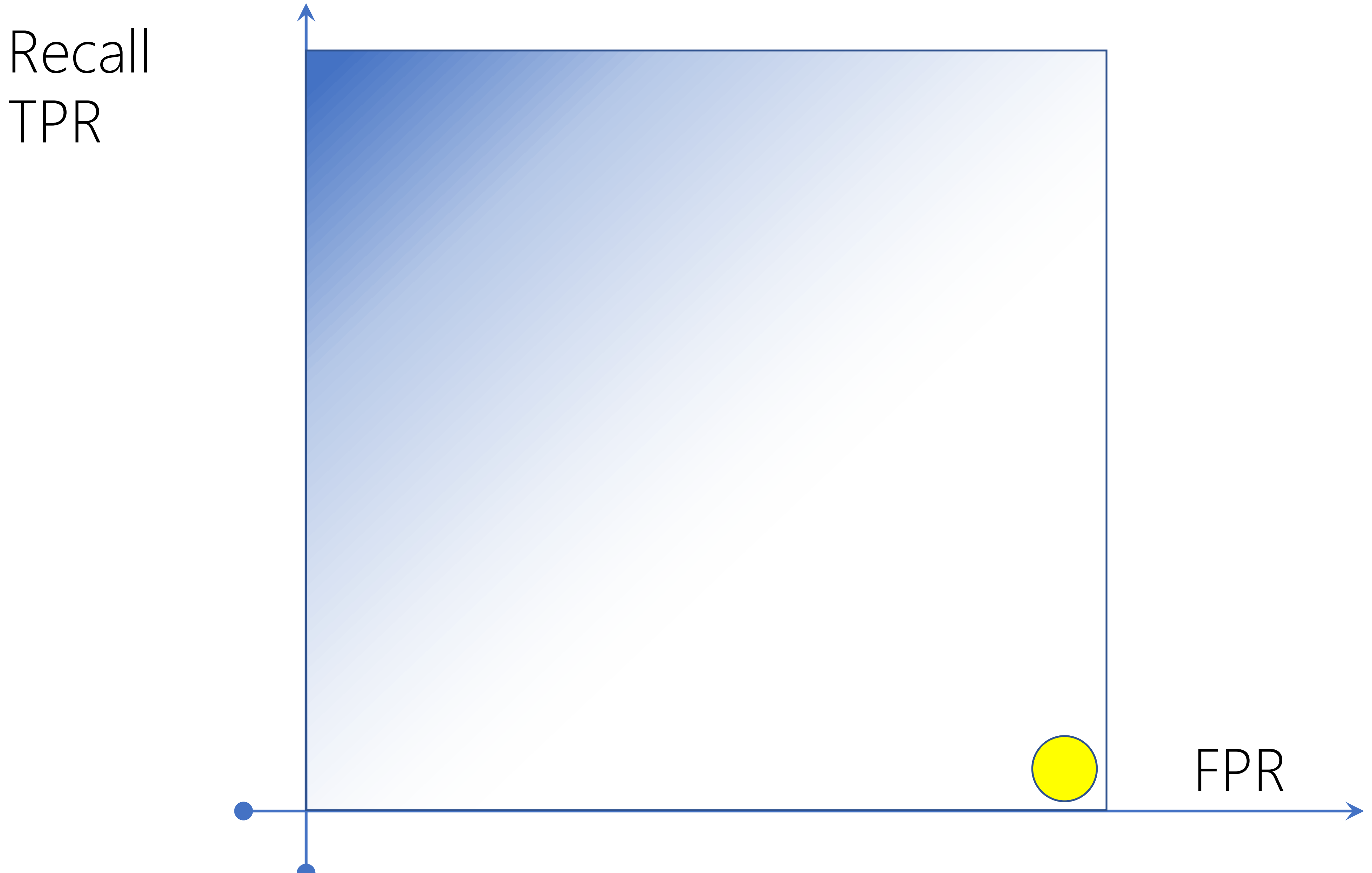
Worst Classifier

	Gold Positive	Gold Negative
Model Positive	0	N-
Model Negative	N+	0

TPR= $\frac{\mathbf{0}}{\mathbf{(}\boldsymbol{N}\mathbf{+)+0}}$ =0.0 ↓

FPR= $\frac{\boldsymbol{N}\mathbf{-}}{\mathbf{0+(}\boldsymbol{N}\mathbf{-)}}$ =1.0 ↑

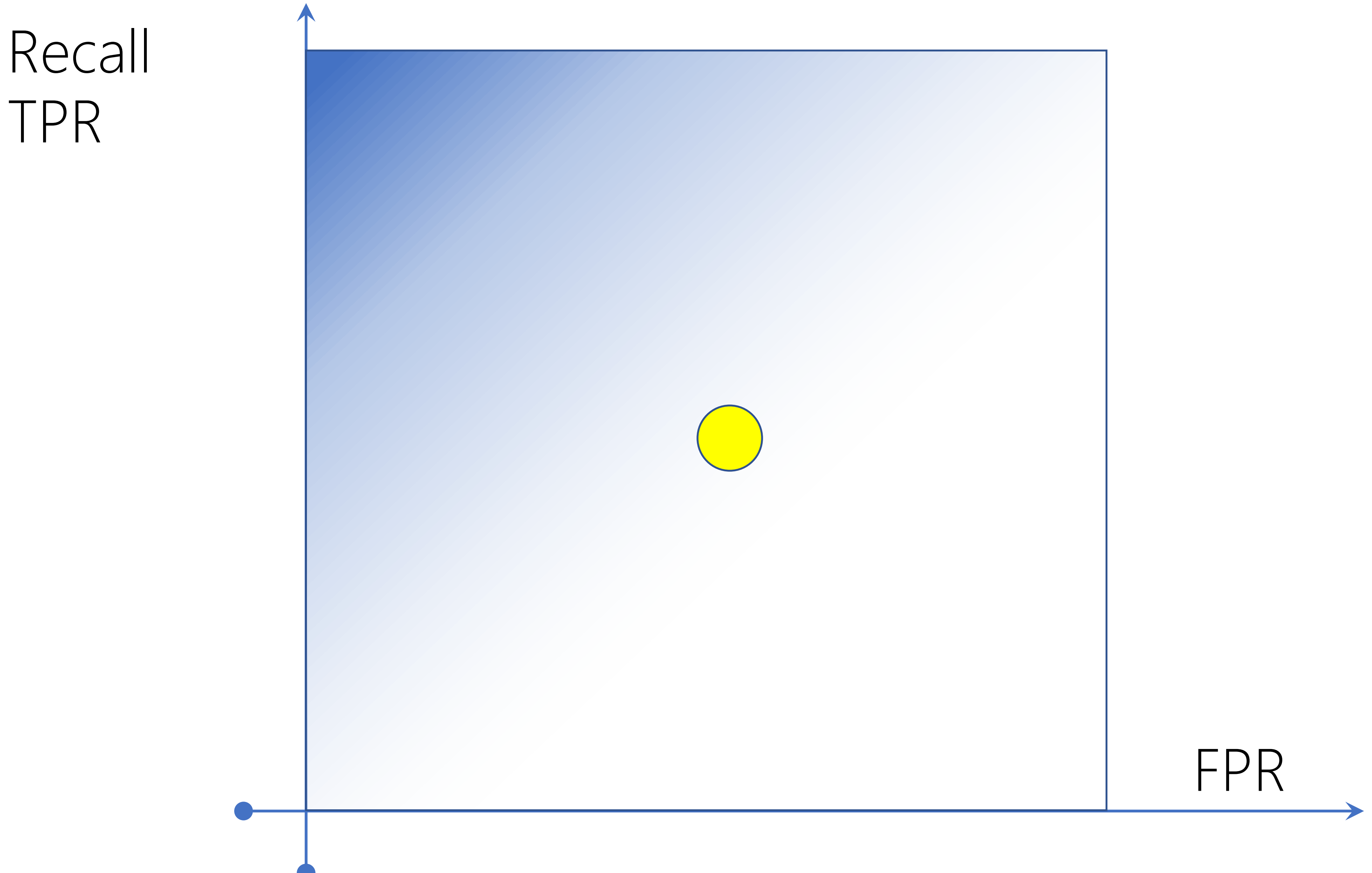
TPR-FPR Curve



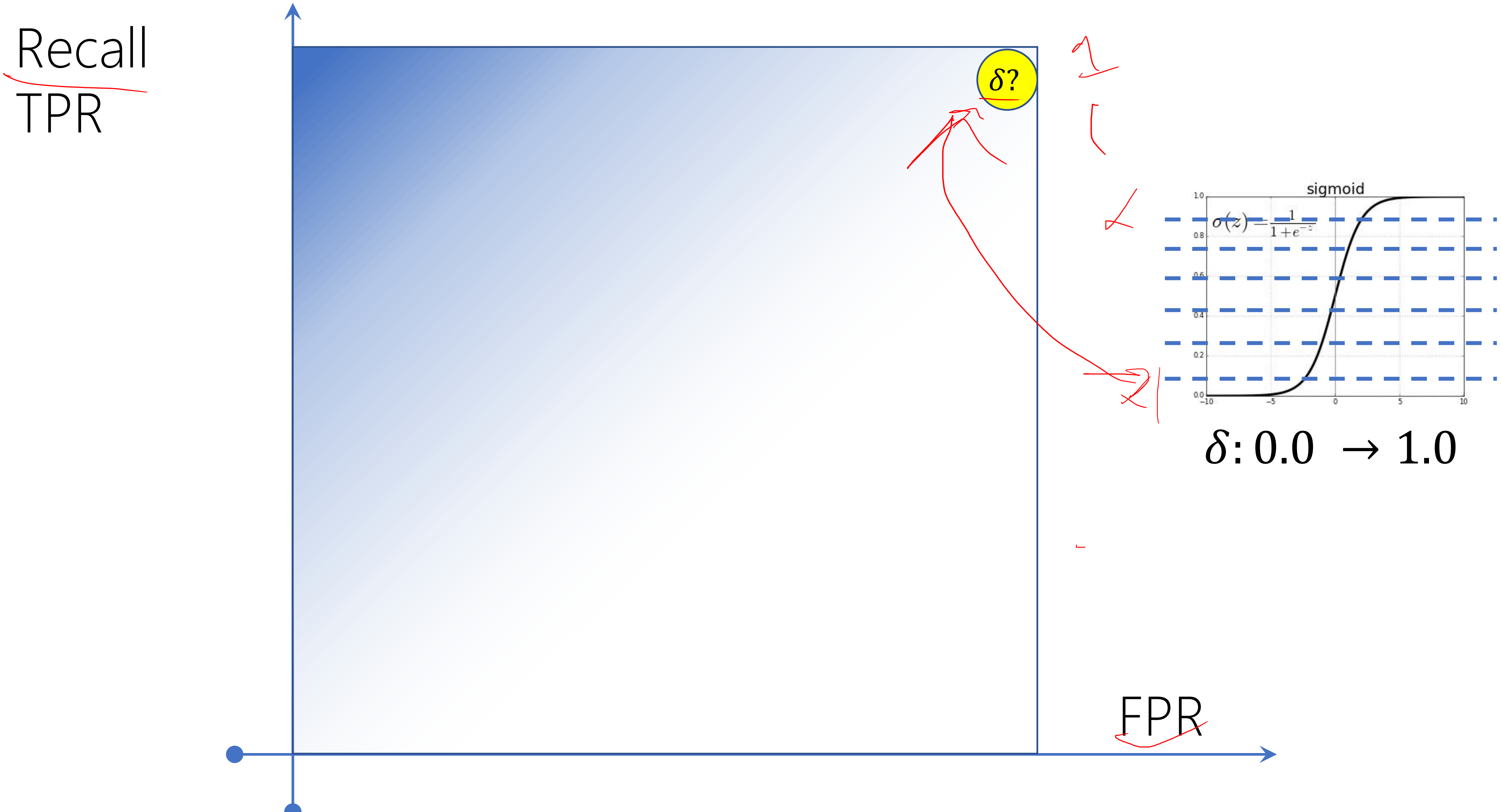
Uniformly Random Classifier

	Gold Positive	Gold Negative
Model Positive	?	?
Model Negative	?	?
	TPR=0.5	FPR=0.5

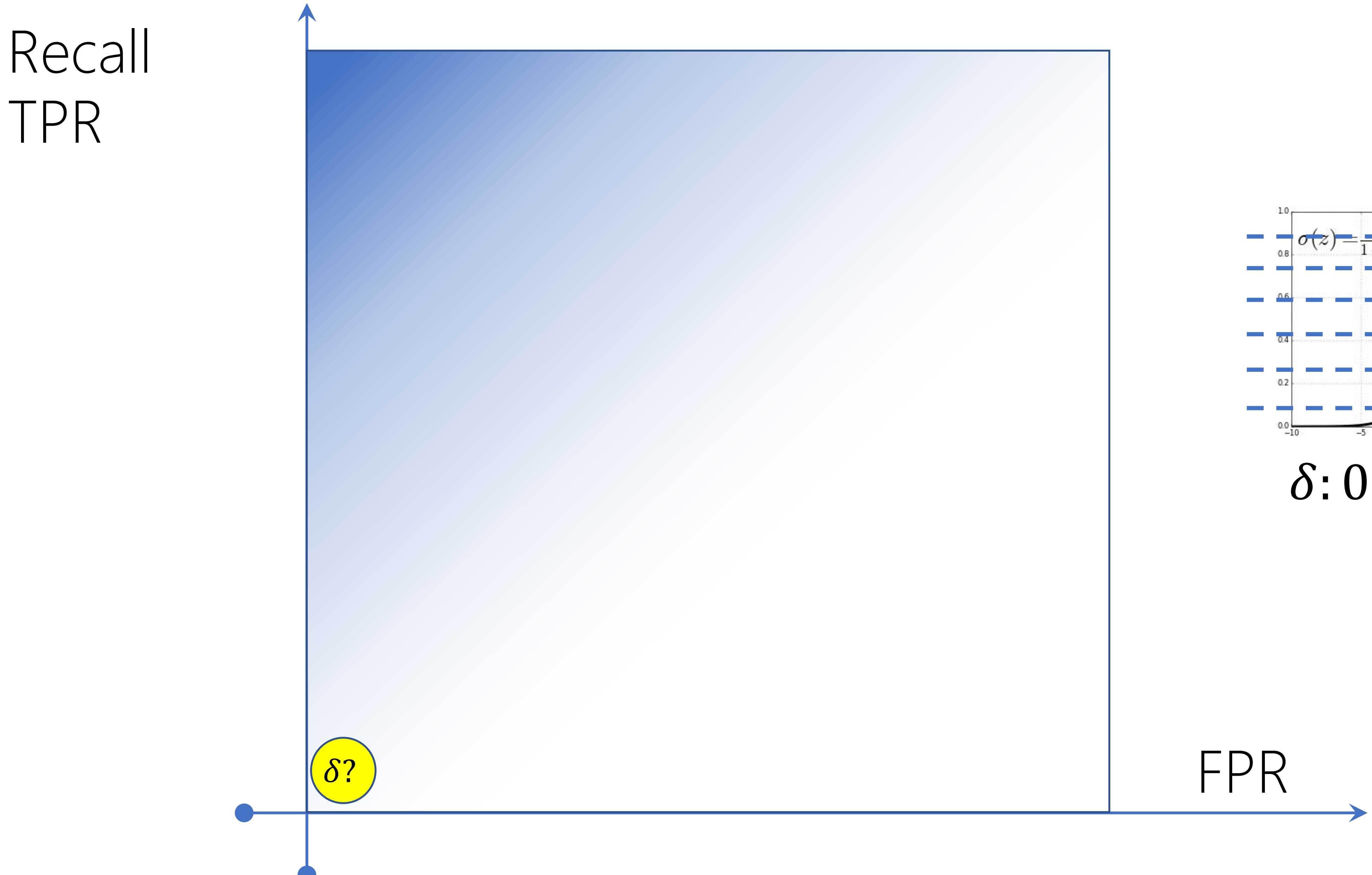
TPR-FPR Curve

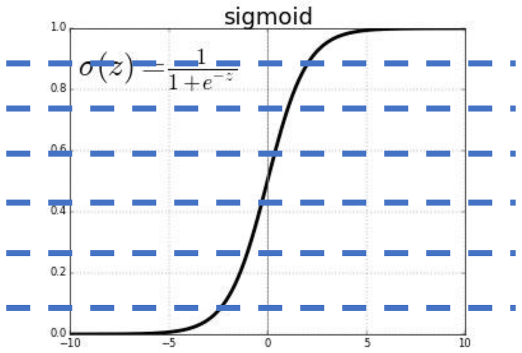


TPR-FPR Curve



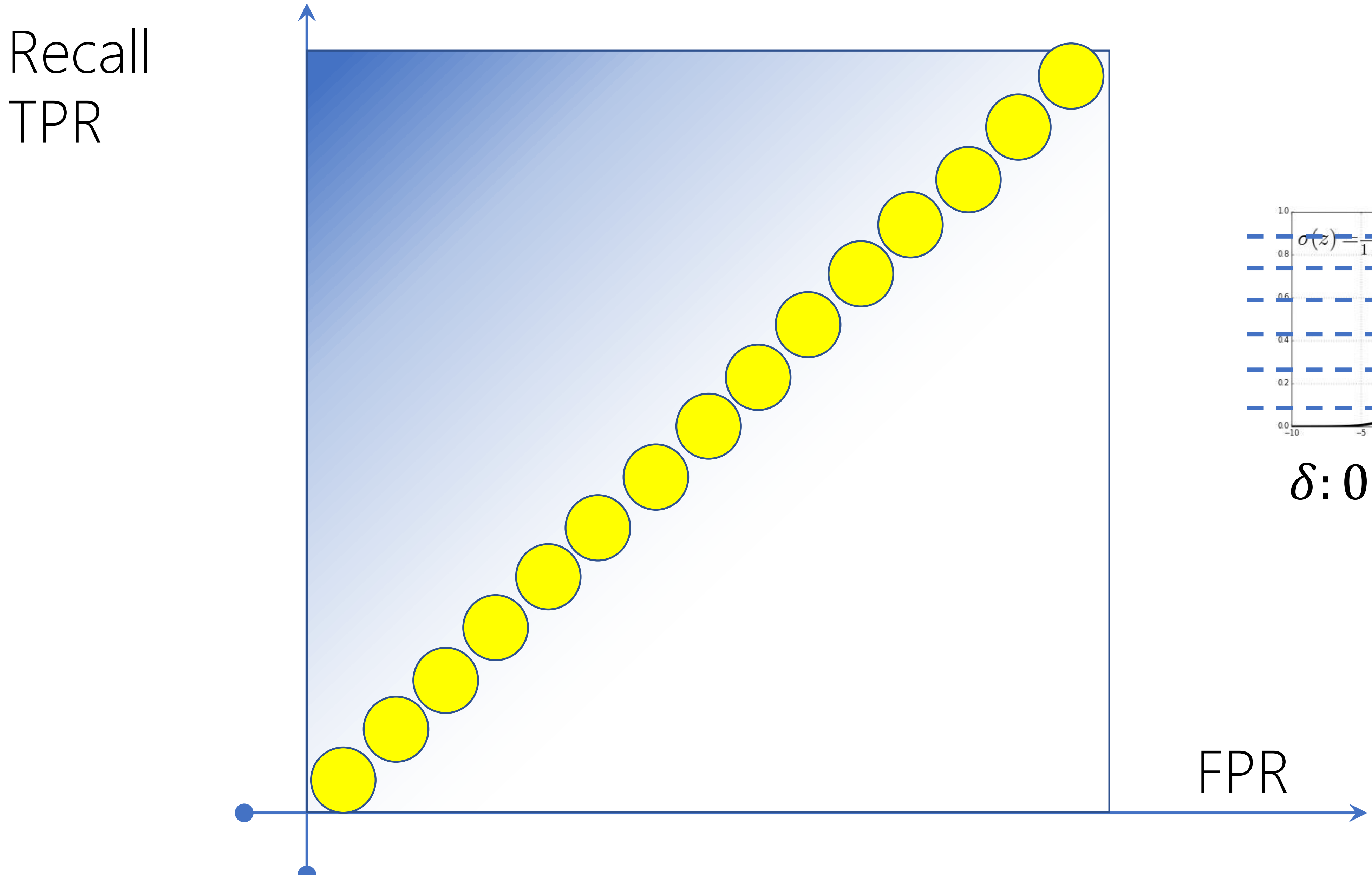
TPR-FPR Curve

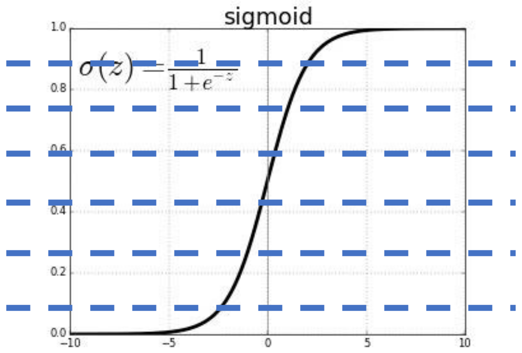




$\delta: 0.0 \rightarrow 1.0$

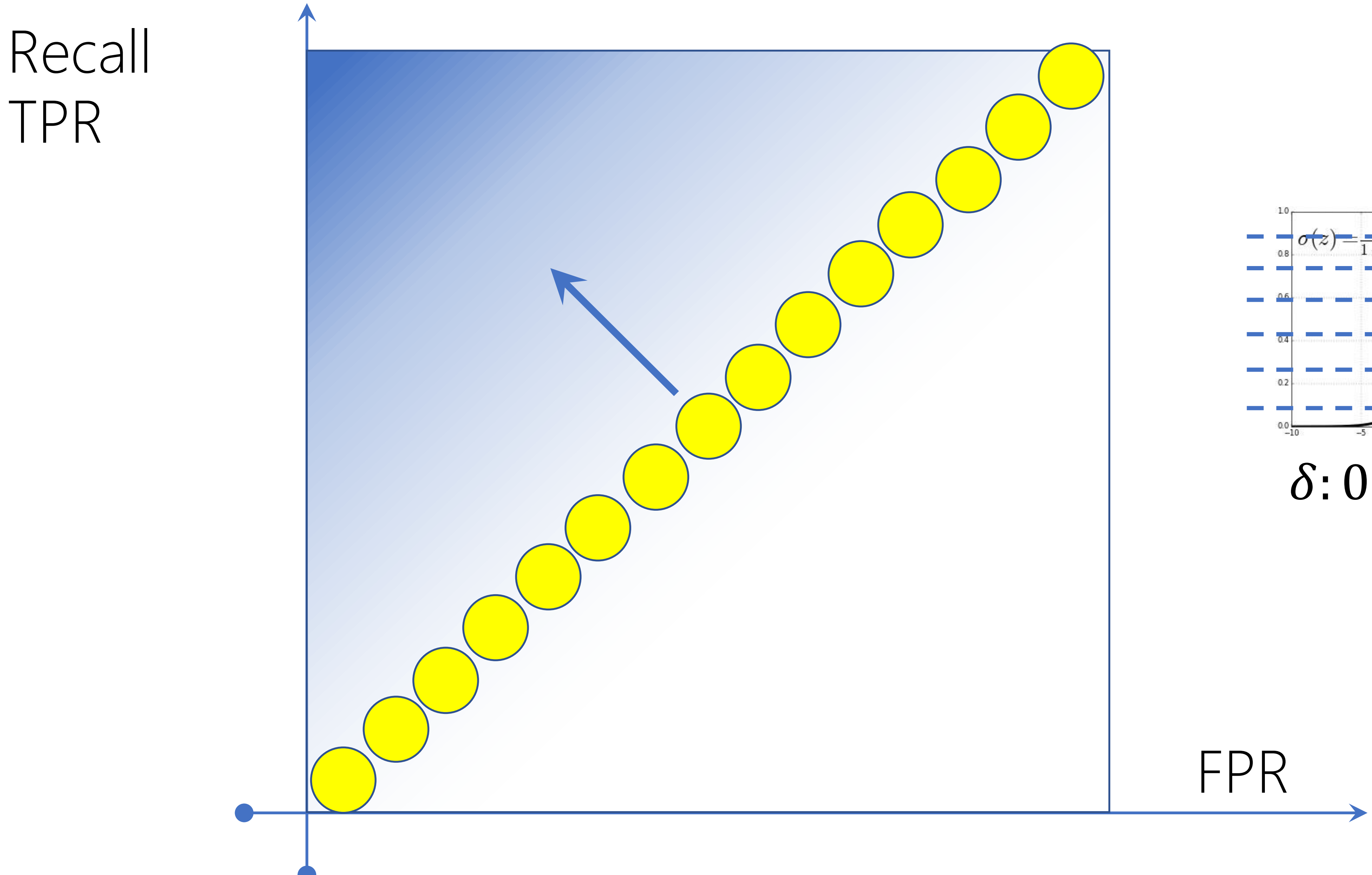
What is this model?

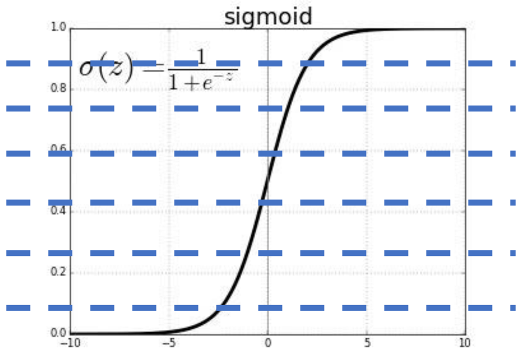




$\delta: 0.0 \rightarrow 1.0$

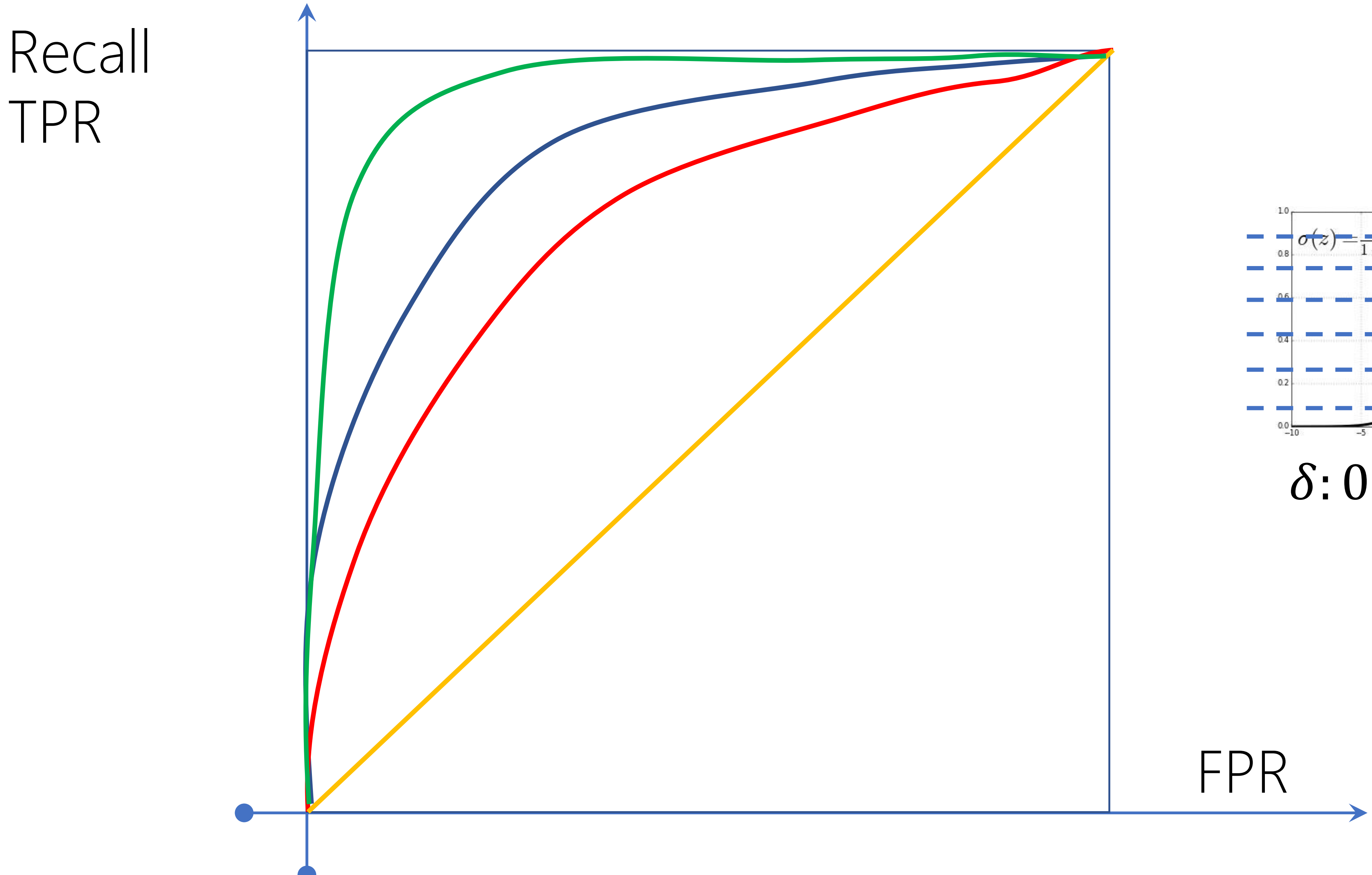
ROC

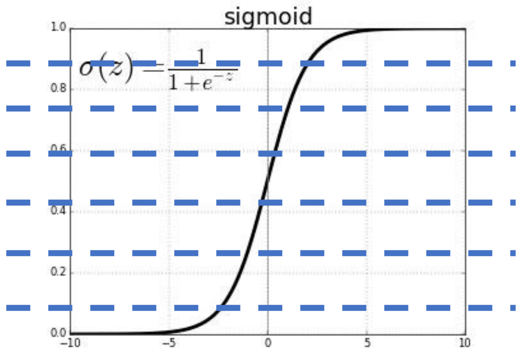




$\delta: 0.0 \rightarrow 1.0$

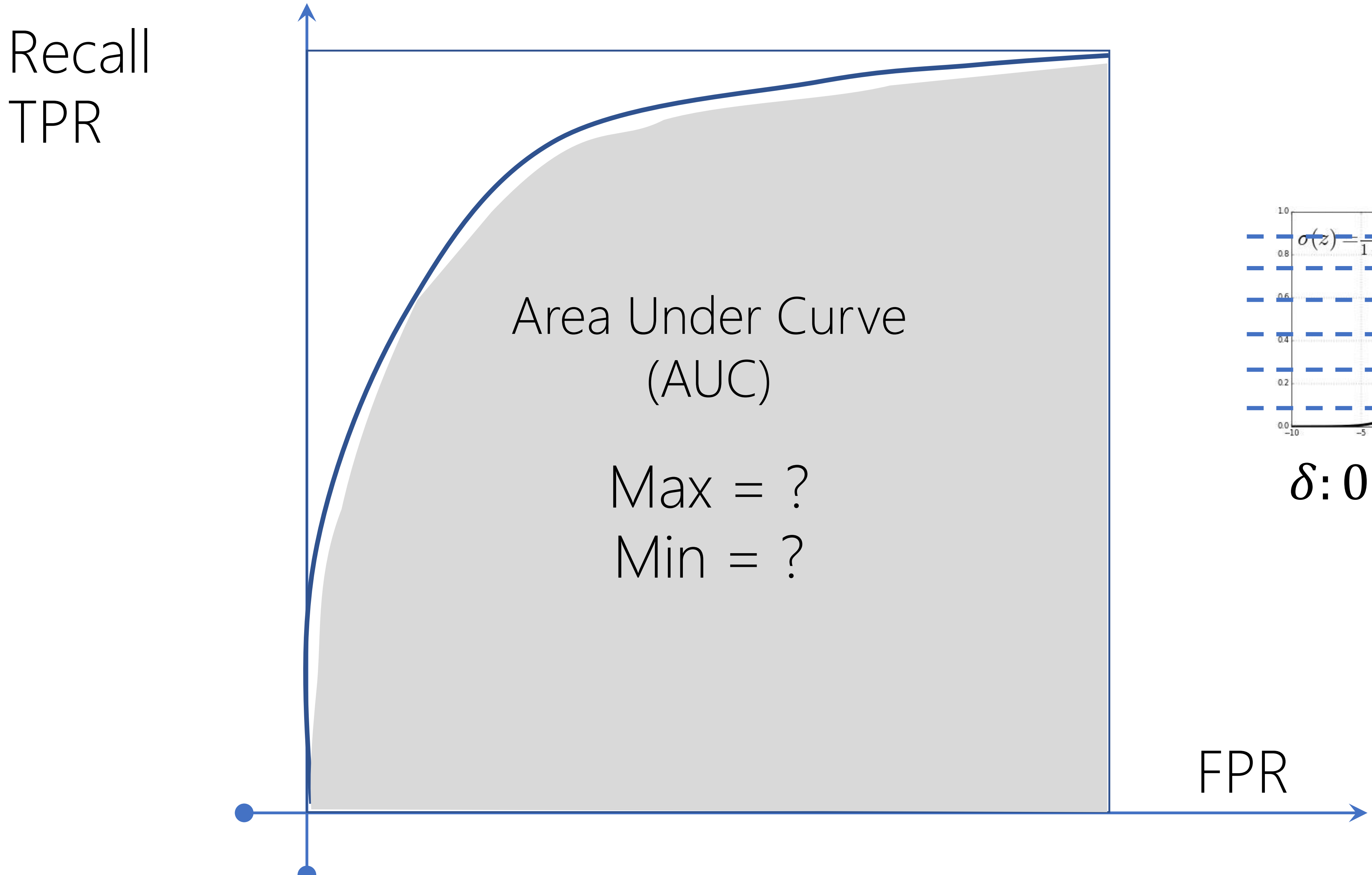
ROC

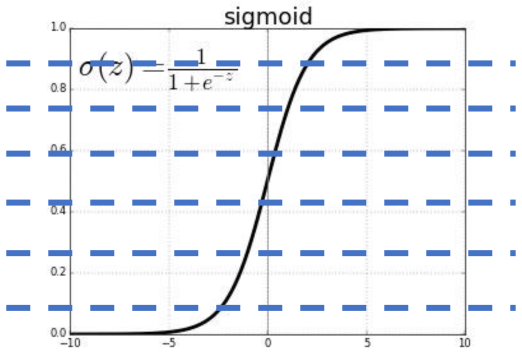




δ: 0.0 → 1.0

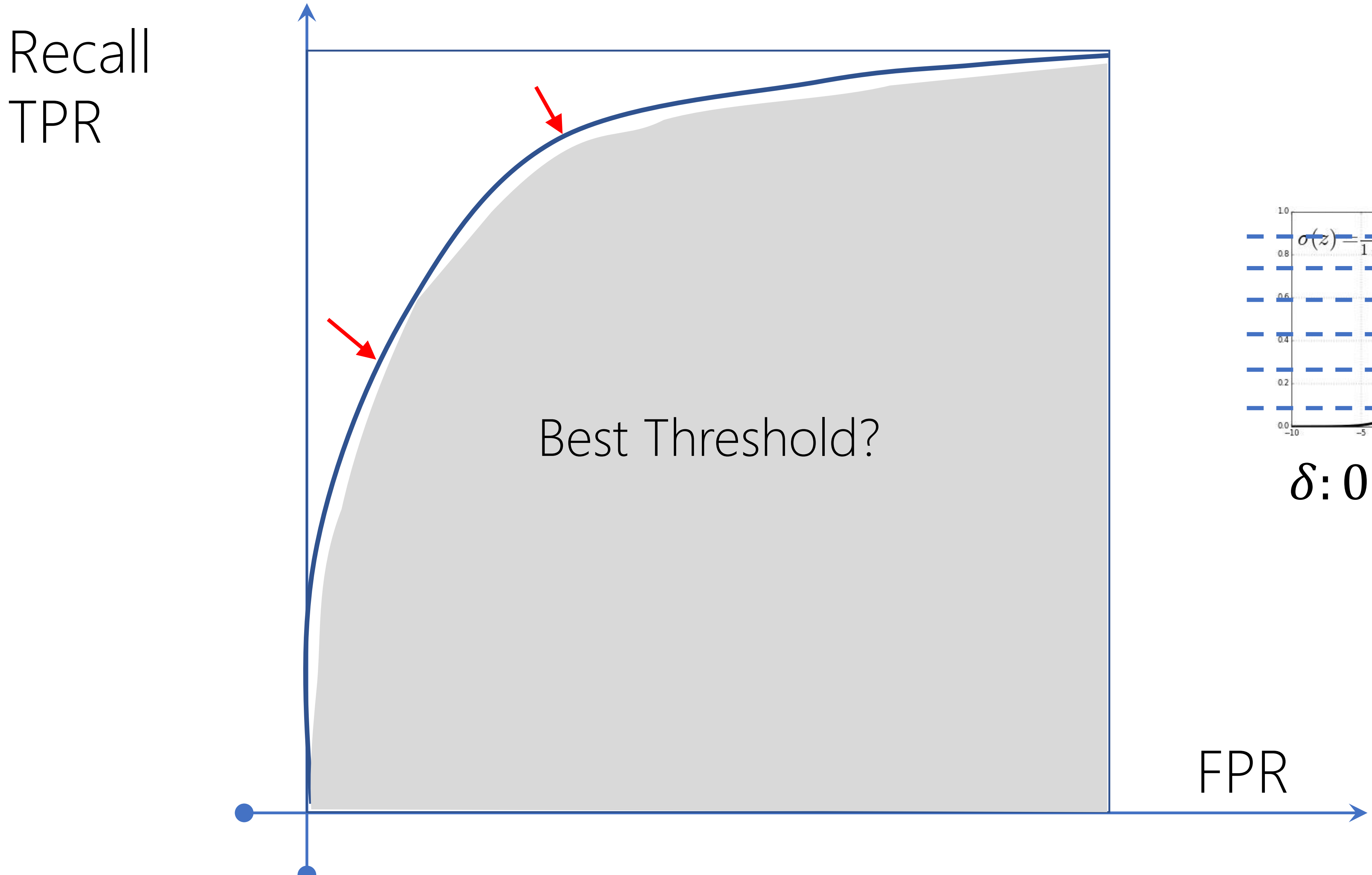
Area Under Curve (AUC): Single Real Point

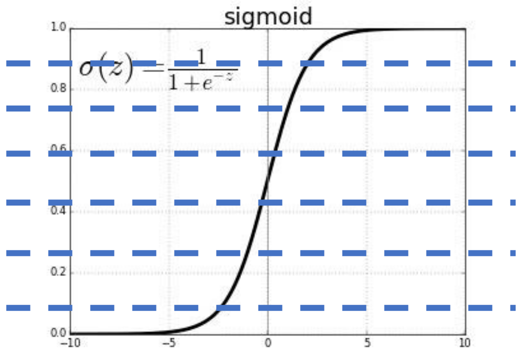




$\delta: 0.0 \rightarrow 1.0$

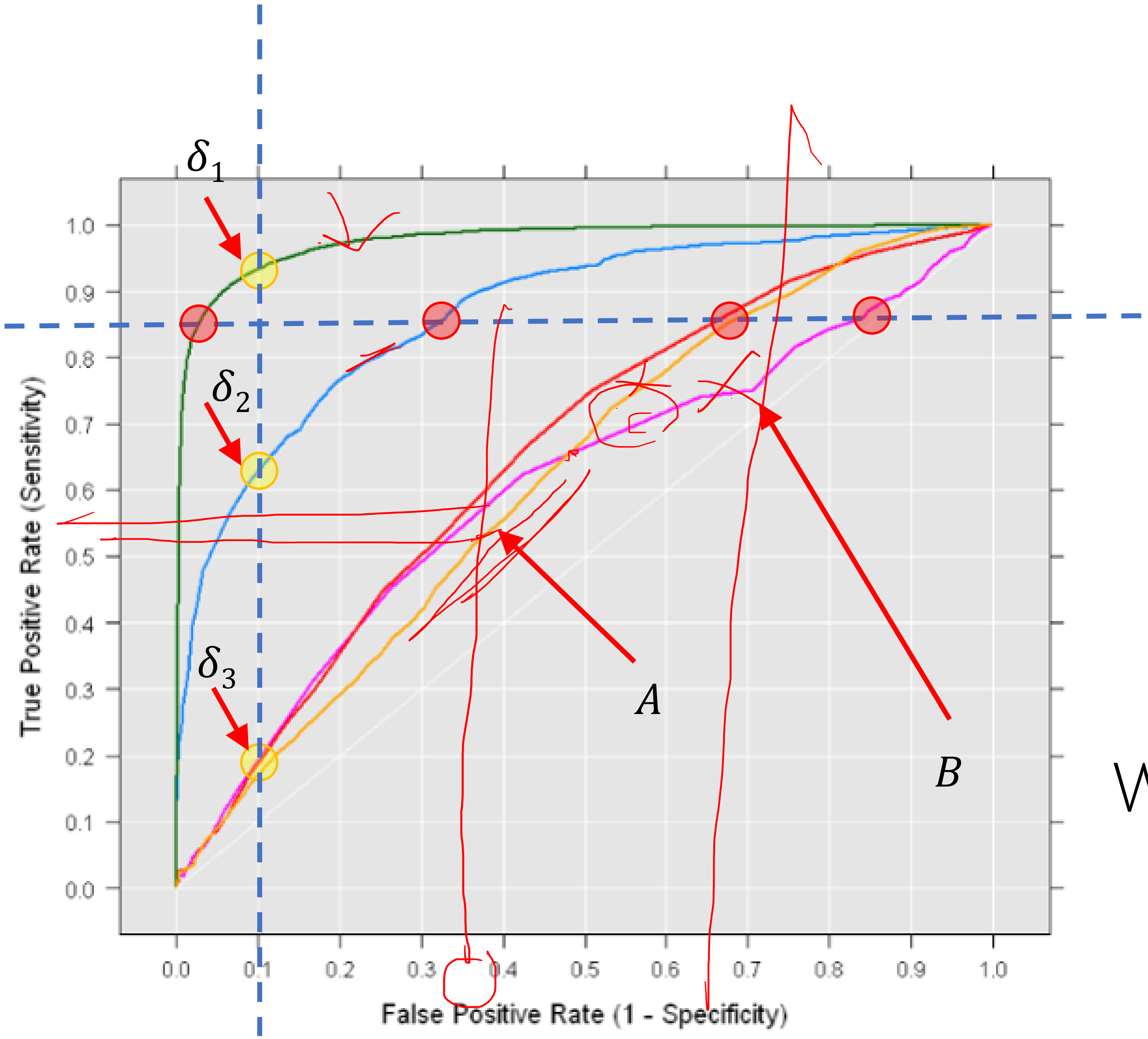
Area Under Curve (AUC): Single Real Point





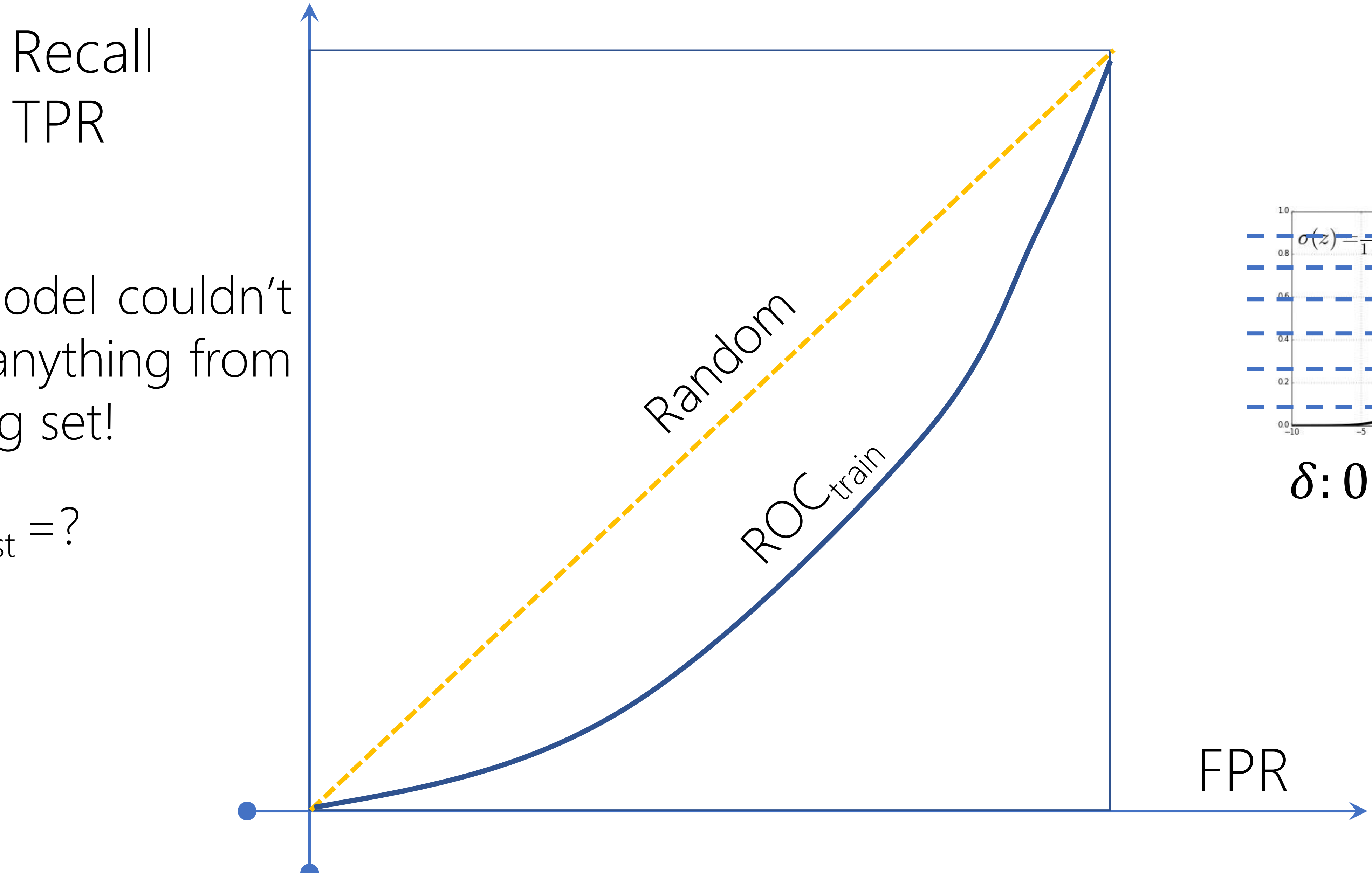
$\delta: 0.0 \rightarrow 1.0$

ROC: Model Comparison



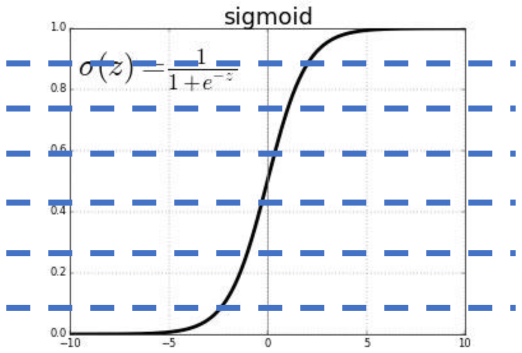
Which one? A or B?

ROC_{train}



The model couldn't learn anything from training set!

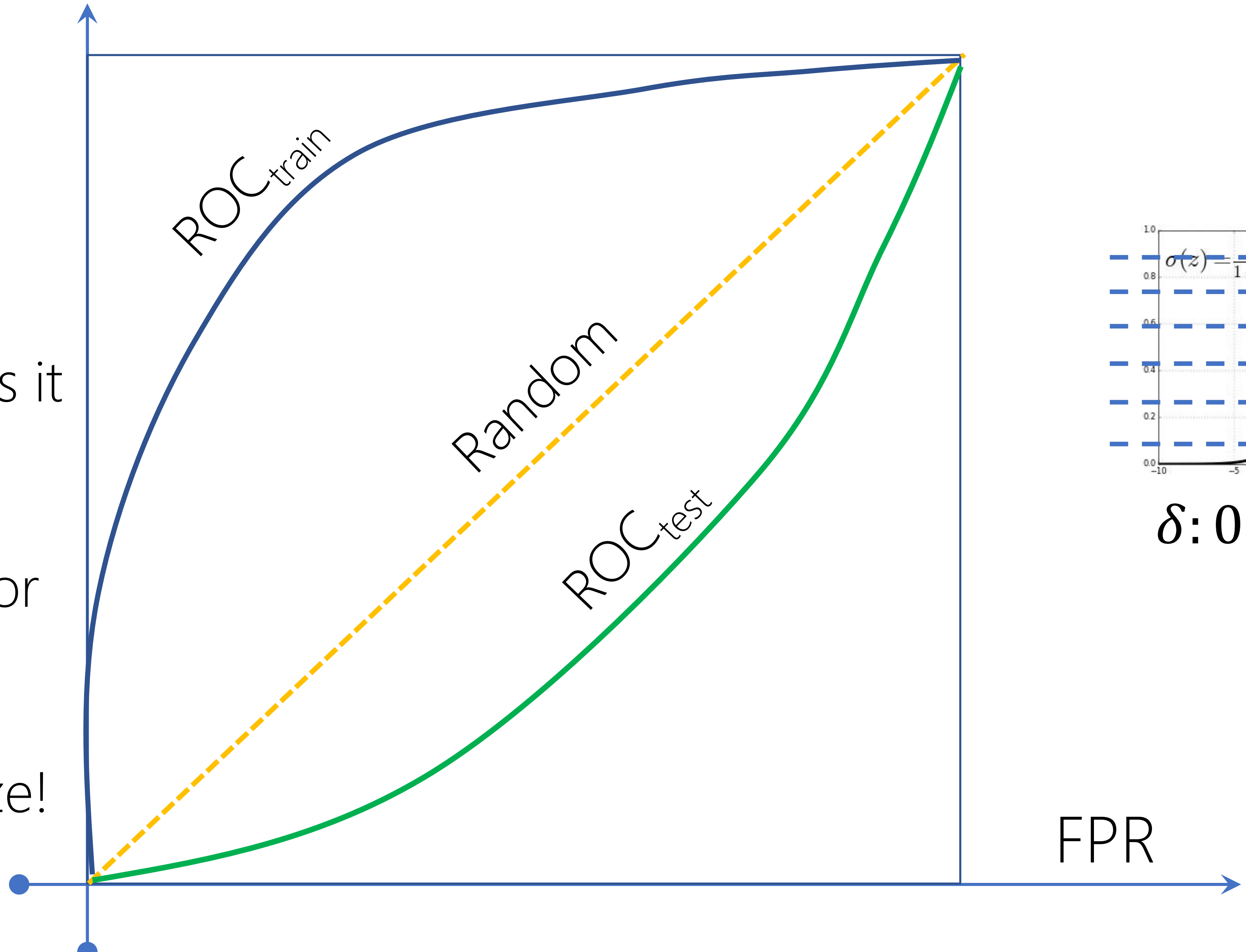
$ROC_{Test} = ?$



$\delta: 0.0 \rightarrow 1.0$

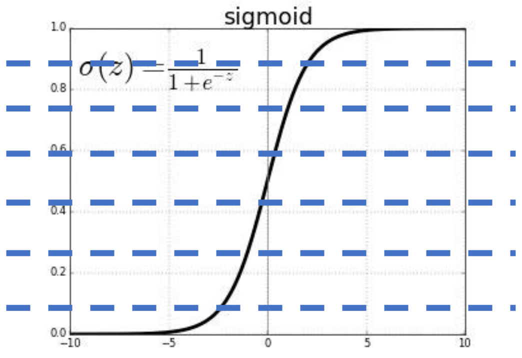
ROC_{test}

Recall
TPR



The model performs well on train set. It means it learnt!

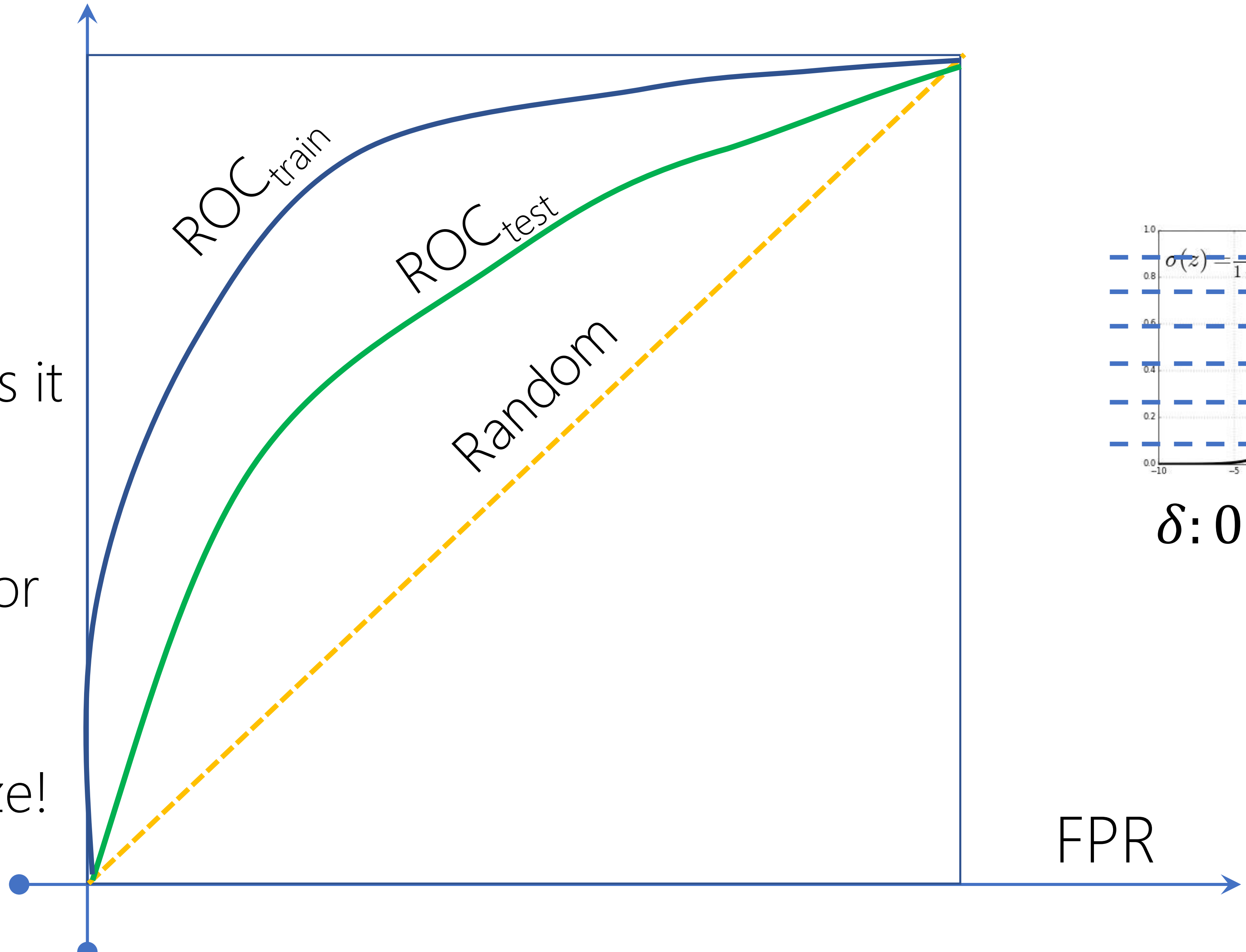
But performs poor in test set. It only know train set. Cannot generalize!



$\delta: 0.0 \rightarrow 1.0$

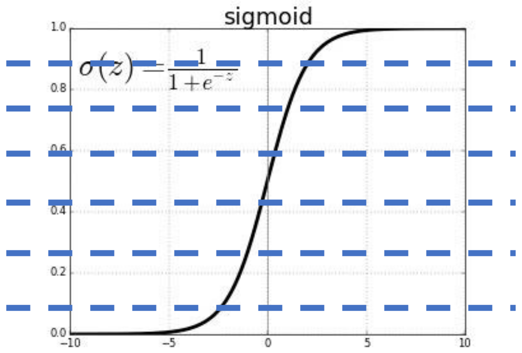
ROC_{test}

Recall
TPR

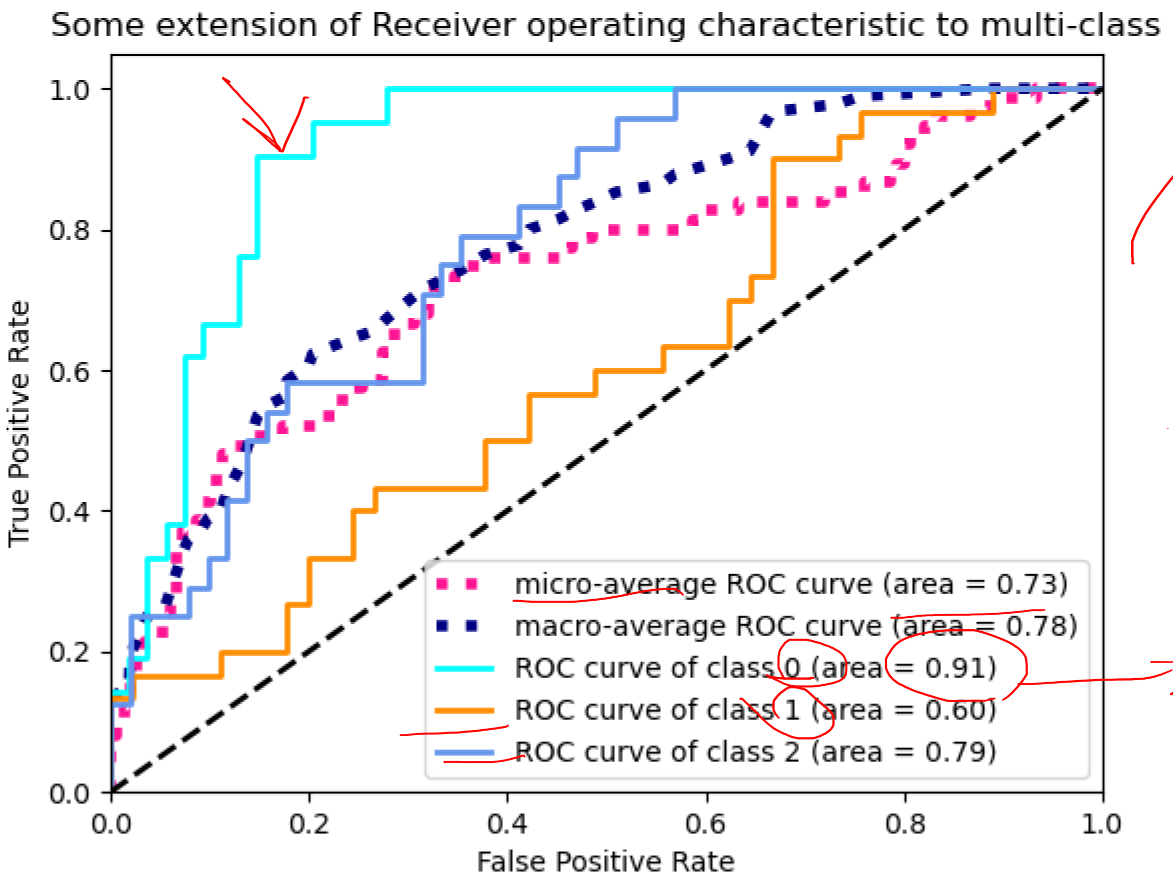


The model performs well on train set. It means it learnt!

But performs poor in test set. It only know train set. Cannot generalize!



$\delta: 0.0 \rightarrow 1.0$



https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

So Far, the LR model and its output

How to input text as X to LR?

X is a vector!

How to map text into vector?

Text = $\{w_1, w_5, w_6\}$

$[1\ 0\ 0\ 0\ 1\ 1\ 0\ 0 \ldots]$

$1(\theta)$

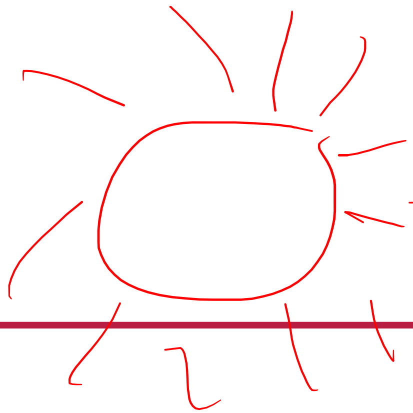
Word
vector space mo

The Treachery of Images

Artist	René Magritte
Year	1929
Medium	Oil on canvas
Movement	Surrealism
Dimensions	60.33 cm × 81.12 cm (23.75 in × 31.94 in)
Location	Los Angeles County Museum of Art[1]

- Phonetics and Phonology
  knowledge about linguistic sounds
- Morphology
  knowledge of the formation and internal structure of words
- Syntax
  knowledge of the structural relationships between words
- Semantics
  knowledge of meaning
- Pragmatics
  knowledge of the relationship of meaning to the goals & intentions of the speaker
- Discourse
  knowledge about linguistic units larger than a single utterance

Task: Engaging in Natural Language Communication



Semiotics: The Science of Symbols

Semantics: Relation between signs and things to which they refer: meaning; sense

Syntactics: Relations among signs in formal structures

Pragmatics: Relation between signs and sign-using agents