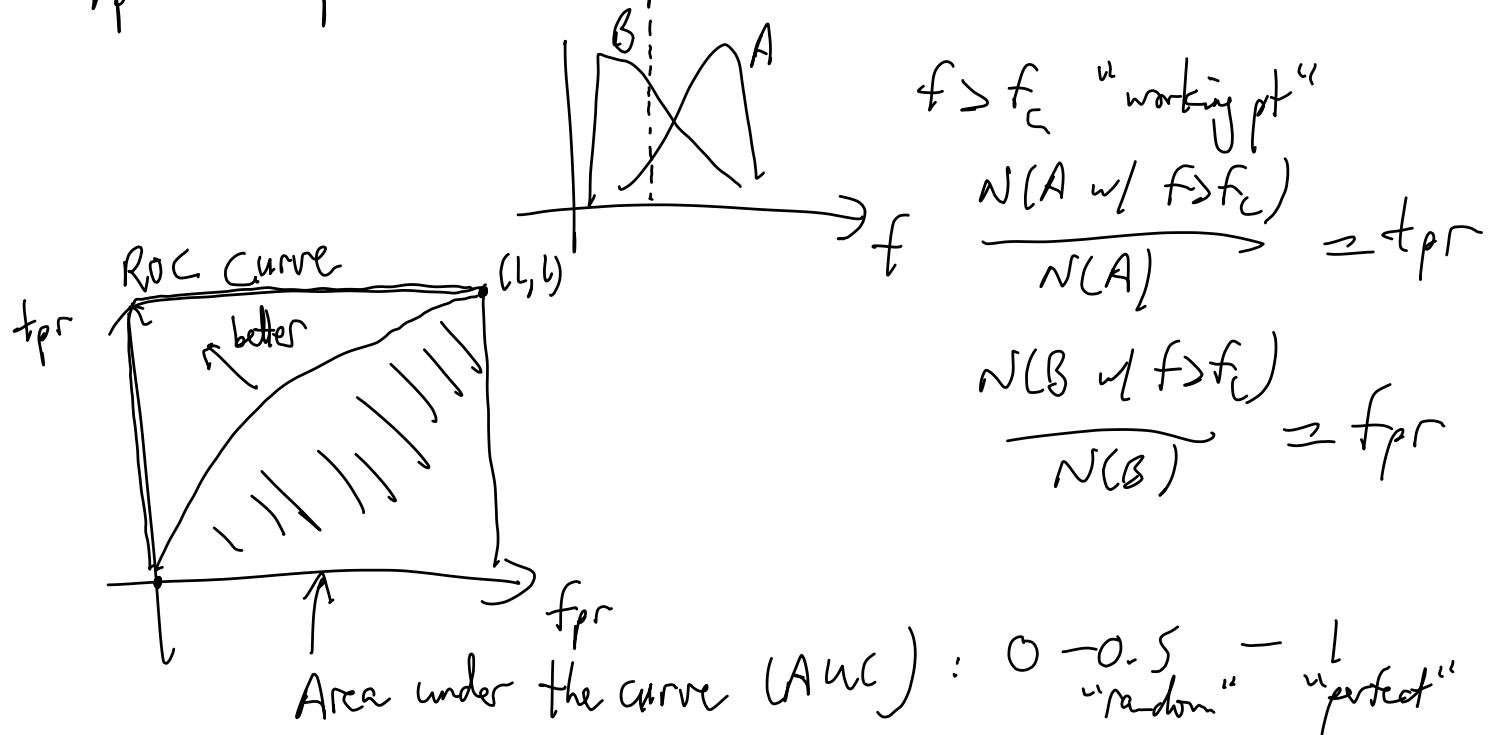


Wednesday, January 27, 2021 10:18 AM

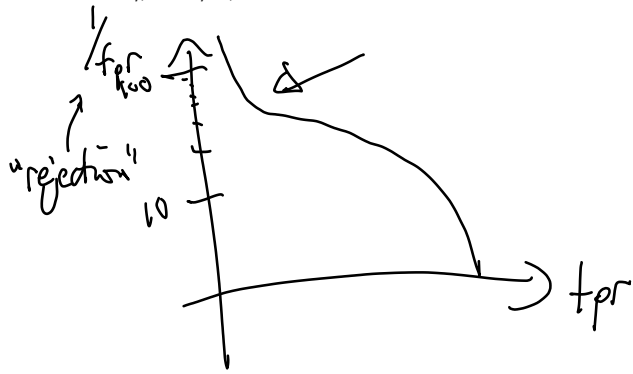
Metrics for binary classification

— log loss itself

— tpr & fpr : take output of classifier $f(x;w)$



Wednesday, January 27, 2021 10:28 AM

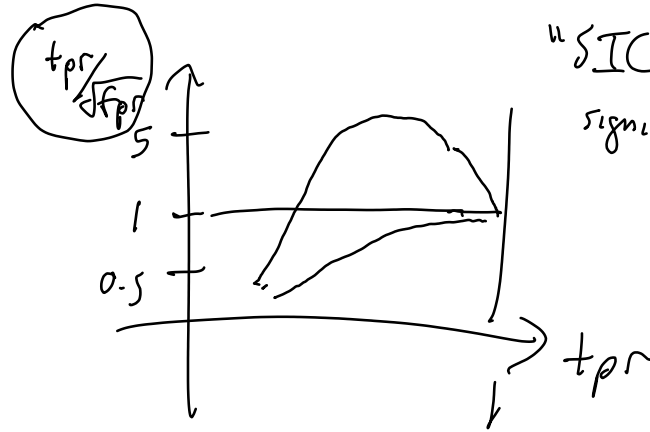


Accuracy: fraction of correct classifications (sig or bs) at the optimal wkt pt

$$Acc = \max \left(\frac{tpr + (1 - fpr)}{2} \right)$$

↓
0.5 "random guessing"
1 "perfect"

↑
for balanced classes



"SIC curve"
significance improvement characteristic

$\frac{S}{\sqrt{B}}$ = common, simple measure of statistical significance
Coming from Poisson distrib.

bs: $B \pm \sqrt{B}$

Before cut on classifier

sig $\frac{S}{\sqrt{B}}$

After " " "

sig $\frac{S \cdot tpr}{\sqrt{B \cdot fpr}}$ indep of S & B in data.



obs: data is $B \pm S$
statistically significant above B?
→ $\frac{S}{\sqrt{B}}$

Wednesday, January 27, 2021 10:43 AM

Why NNs?

Universal Approx. Thm :

"An infinitely wide, single layer NN can approx any continuous fn to arbitrary precision."

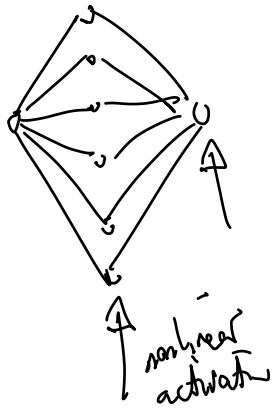
"An infinitely deep, fixed width NN can also yada yada..."

"Any nonpolynomial activation fn"

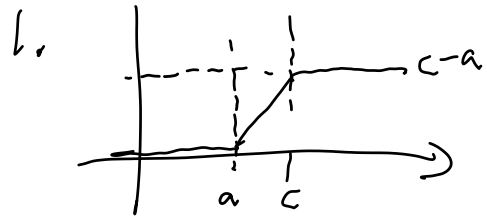
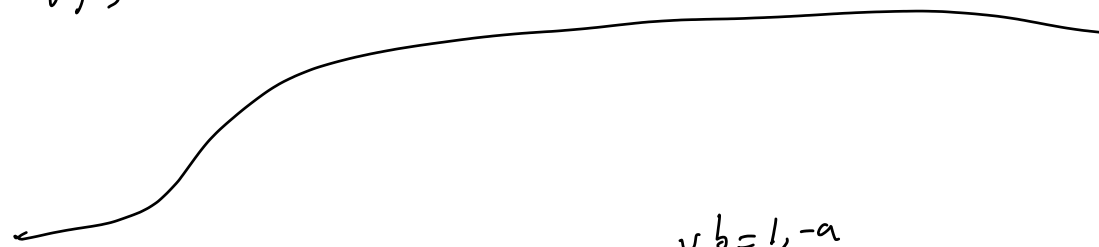
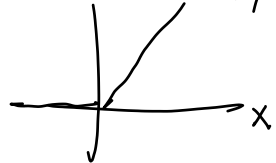
sigmoid
↑
(original)

(subsequent)

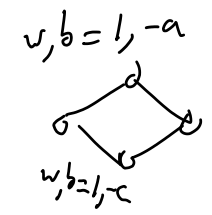
(subsequent)



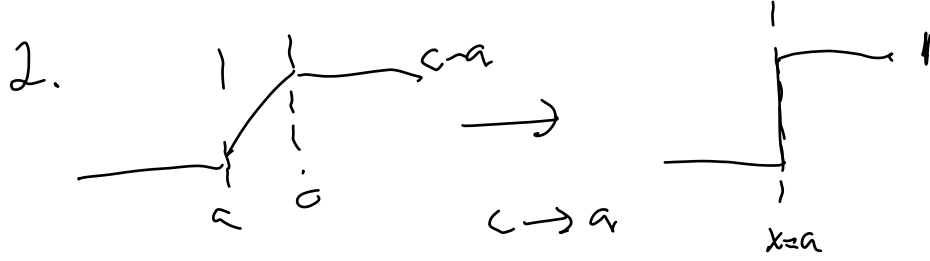
"proof" for 1d fn. $f(x)$
ReLU activation
 $A(x) = \max(0, x)$



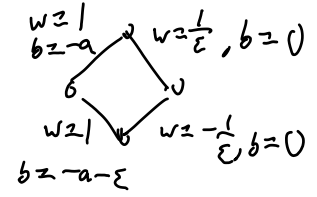
$$\max(0, x-a) - \max(0, x-c)$$



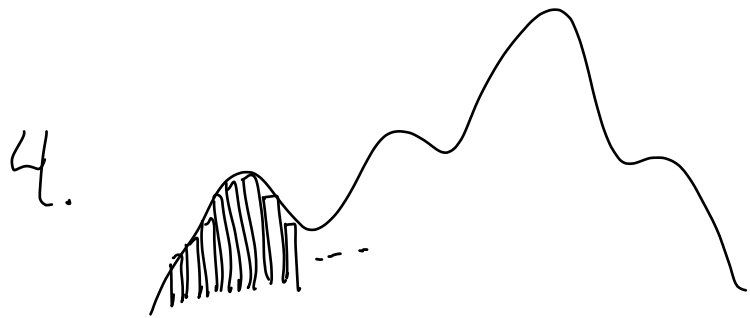
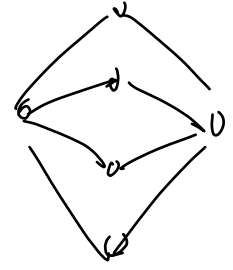
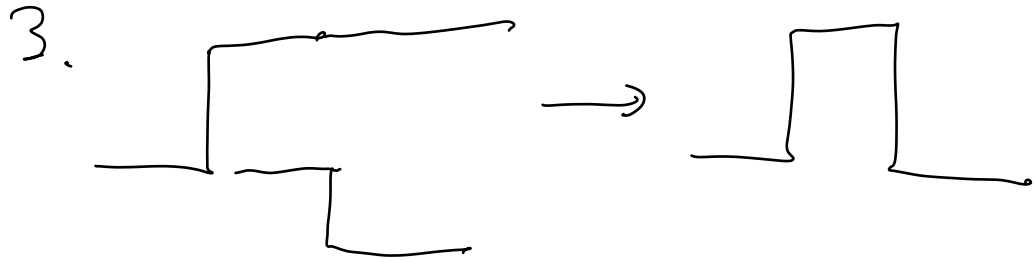
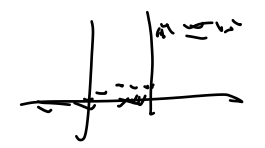
Wednesday, January 27, 2021 10:55 AM



$$\frac{1}{\epsilon} \left(\max(0, x-a) - \max(0, x-a-\epsilon) \right)$$



Ex: try this out on "detect"



Wednesday, January 27, 2021 11:05 AM

MNIST

"mouse model" for ML

70,000 handwritten digits

28x28 grayscale images

(pixels take values
integers btw 0-255)