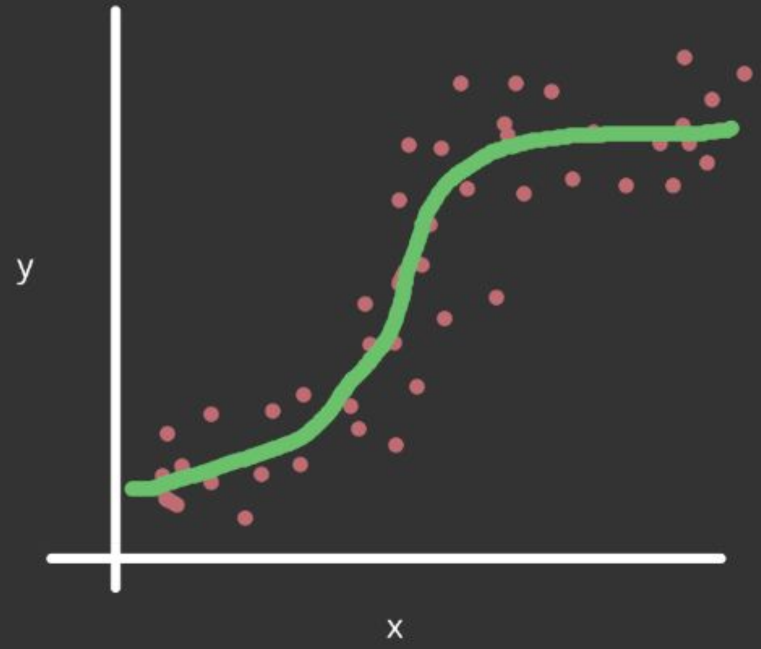
L05

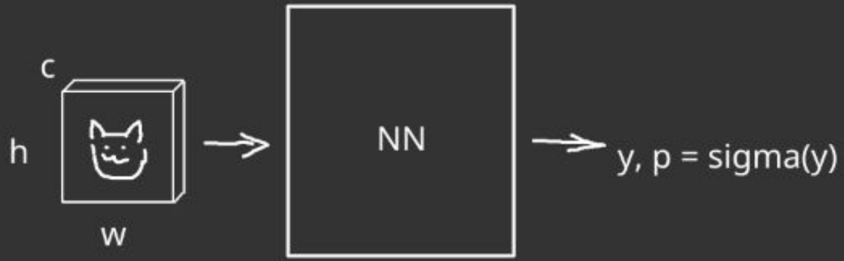
Classification

Regression



Binary Classification

Cat or Dog?



Sigmoid is activation. But what is the loss? MSE?

$$\sigma(y) = \frac{1}{1+e^{-y}}$$

$$\frac{d\sigma}{dy} = \sigma(1-\sigma)$$

$$MSE = (\sigma(y) - t)^2$$

$$MSE' = 2(\sigma(y) - t)\sigma' = 2(\sigma(y) - t)\sigma(1-\sigma)$$

Sigmoid is activation. But what is the loss? MSE?

$$\sigma(y) = \frac{1}{1+e^{-y}}$$

$$\frac{d\sigma}{dy} = \sigma(1-\sigma)$$

$$MSE = (\sigma(y) - t)^2$$

$$MSE' = 2(\sigma(y) - t)\sigma' = 2(\sigma(y) - t)\sigma(1 - \sigma)$$

- independent on error
- can be easily become 0
- lead to NN paralysis

Sigmoid is activation. But what is the loss? BCE?

$$\sigma(y) = \frac{1}{1+e^{-y}} = p$$

$$\frac{d\sigma}{dy} = \sigma(1-\sigma)$$

$$BCE(p,t) = -t\log(p) - (1-t)\log(1-p)$$

Check if there are no log(0) or log of negative number

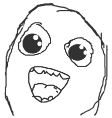
Sigmoid is activation. But what is the loss? BCE?

$$\sigma(y) = \frac{1}{1+e^{-y}} = p \qquad \frac{d\sigma}{dy} = \sigma(1-\sigma)$$

$$BCE(p,t) = -t\log(p) - (1-t)\log(1-p)$$

$$\frac{d(BCE)}{dy} = \frac{d(BCE)}{dp}\frac{dp}{dy} = -\frac{t}{\sigma}\sigma' + \frac{1-t}{1-\sigma}\sigma' =$$

$$= -t(1-\sigma) + (1-t)\sigma = \sigma - t$$

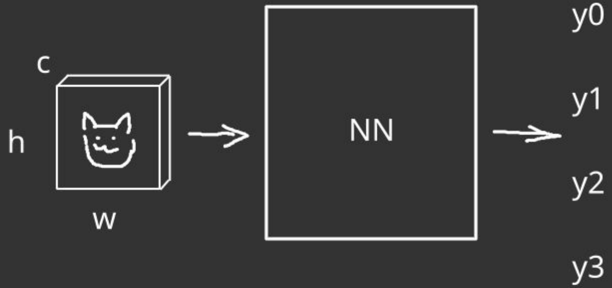


BCE is Good for

- Probability estimation
- For estimation of any numbers in (0, 1) range

Multiclass Classification

Cat, Dog, Pig, or Plane?



p(y0) + p(y1) + p(y2) + p(y3) = 1

Ref: link

Multiclass Classification. Activation Function. Softmax

Must watch: https://www.youtube.com/watch?v=8ps_JEW42xs

$$SM_i(\vec{y}) = \frac{e^{y_i}}{\sum_{j=1}^{N} e^{y_j}}$$

$$SM_2([2,3,5,0]) = \frac{e^5}{e^2 + e^3 + e^5 + e^0} = 0.84$$

$$SM([2,3,5,0]) = [0.04, 0.11, 0.84, 0.01]$$

Softmax Properties

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$SM_i(\overrightarrow{y}) = \frac{e^{y_i}}{\sum_{j=1}^{N} e^{y_j}}$$

$$0 < SM < 1$$

$$\sum_{i}^{N} SM_i = 1$$

$$\frac{d(SM_i)}{dy_c} = \frac{-e^{y_i} e^{y_c}}{(\sum_{j=1}^{N} e^{y_j})^2} = -SM_i \cdot SM_c, c \neq y$$

Softmax Properties

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$SM_i(\overrightarrow{y}) = \frac{e^{y_i}}{\sum_{j=1}^{N} e^{y_j}}$$

$$\frac{d(SM_i)}{dy_i} = \frac{e^{y_i}\sum_{j=1}^{N} e^{y_j} - e^{y_i}e^{y_i}}{(\sum_{j=1}^{N} e^{y_j})(\sum_{j=1}^{N} e^{y_j})} = SM_i \cdot (1 - SM_i)$$

Loss Function for Softmax: Cross Entropy

$$p_c = SM_c(\overrightarrow{y})$$

$$CE(p, t) = -\sum_{c=1}^{N} t_c \log p_c$$

Task: take 2 distributions and calculate *CE*

Loss Function for Softmax: Cross Entropy. Is It Good?

$$p_c = SM_c(\overrightarrow{y})$$

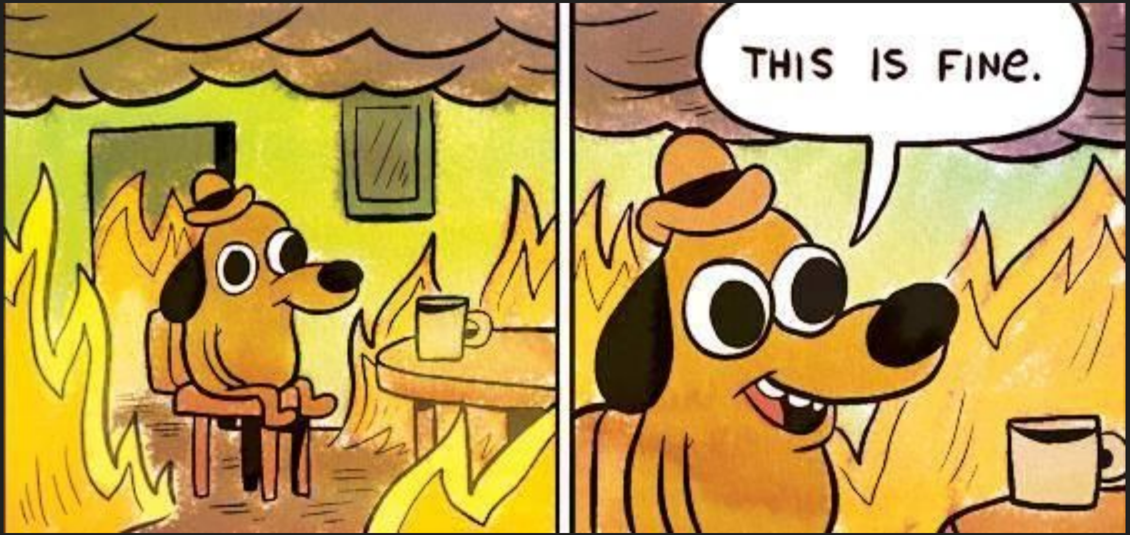
$$CE(p,t) = -\sum_{c=1}^{N} t_c \log p_c$$

$$\boxed{\frac{d(CE)}{dy_i}} = \frac{-\sum_{c=1}^{N} t_c \log p_c}{dy_i} =$$

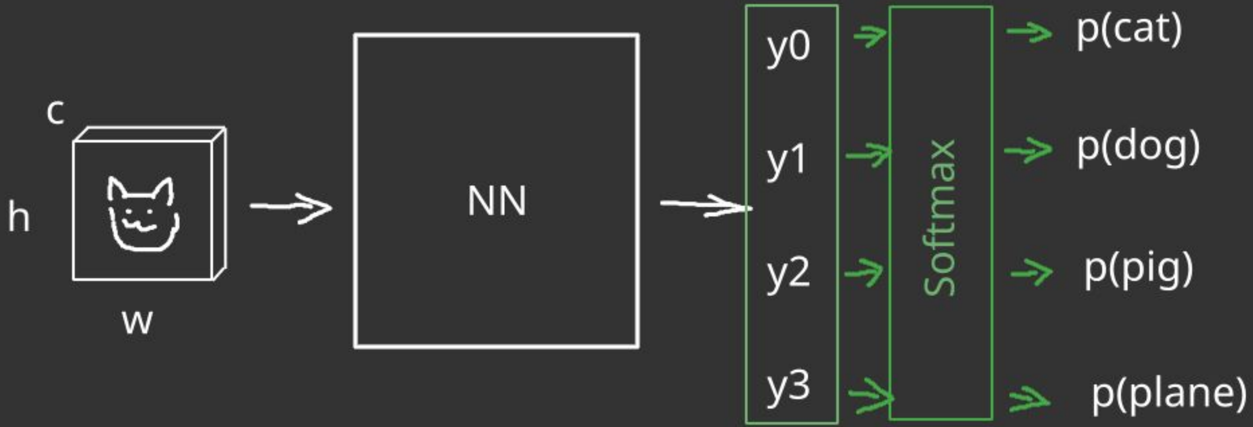
$$= -\frac{t_i}{p_i}\frac{d(SM_i)}{dy_i} - \sum_{c \neq i} \frac{t_c}{p_c}\frac{d(SM_c)}{dy_i} =$$

$$= -\frac{t_i}{p_i} p_i(1-p_i) + \sum_{c \neq i} t_c \frac{1}{p_c} p_i p_c = -t_i(1-p_i) + \sum_{c \neq i} t_c p_i = \boxed{-t_i + p_i}$$

Let's take a break and sip tea



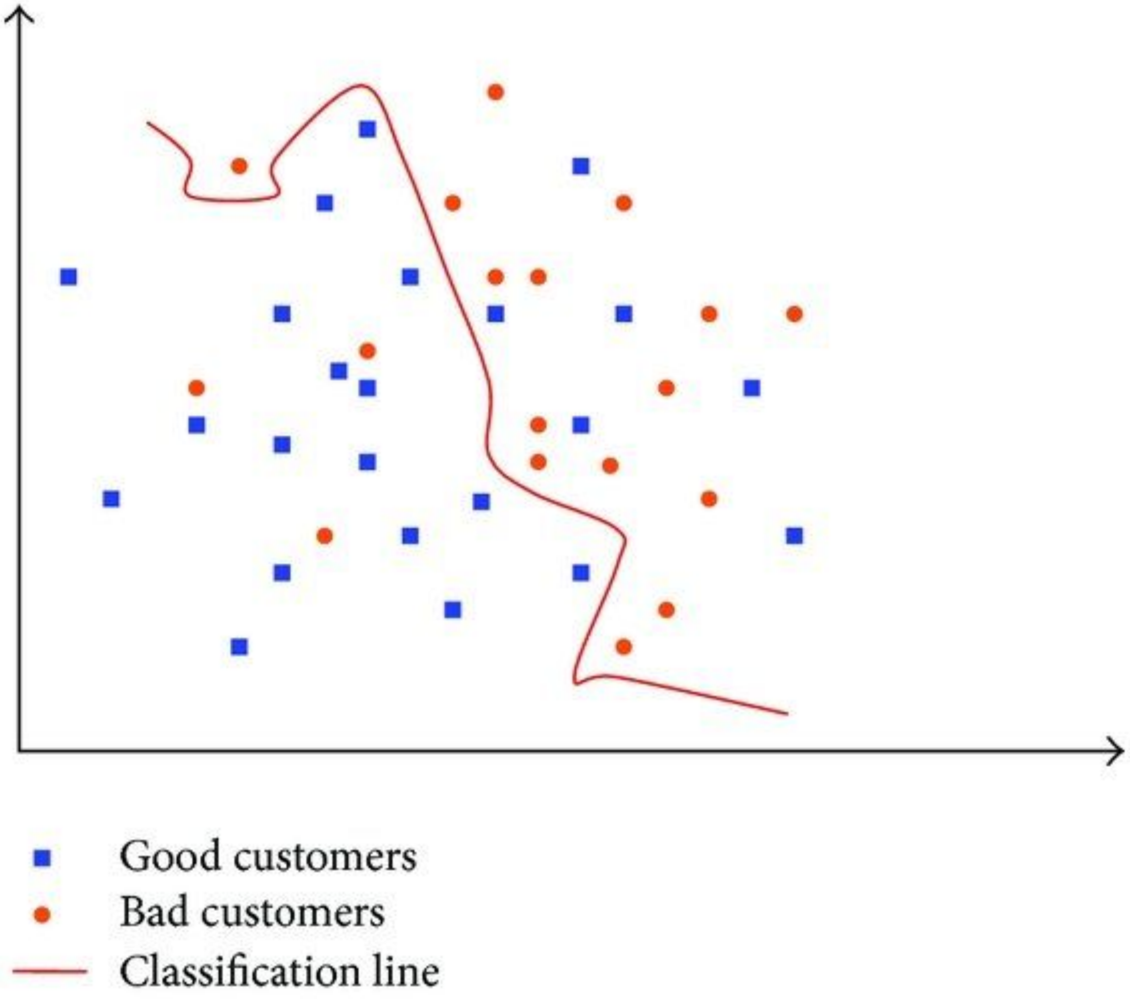
A lot of math. What about multiclass classification?



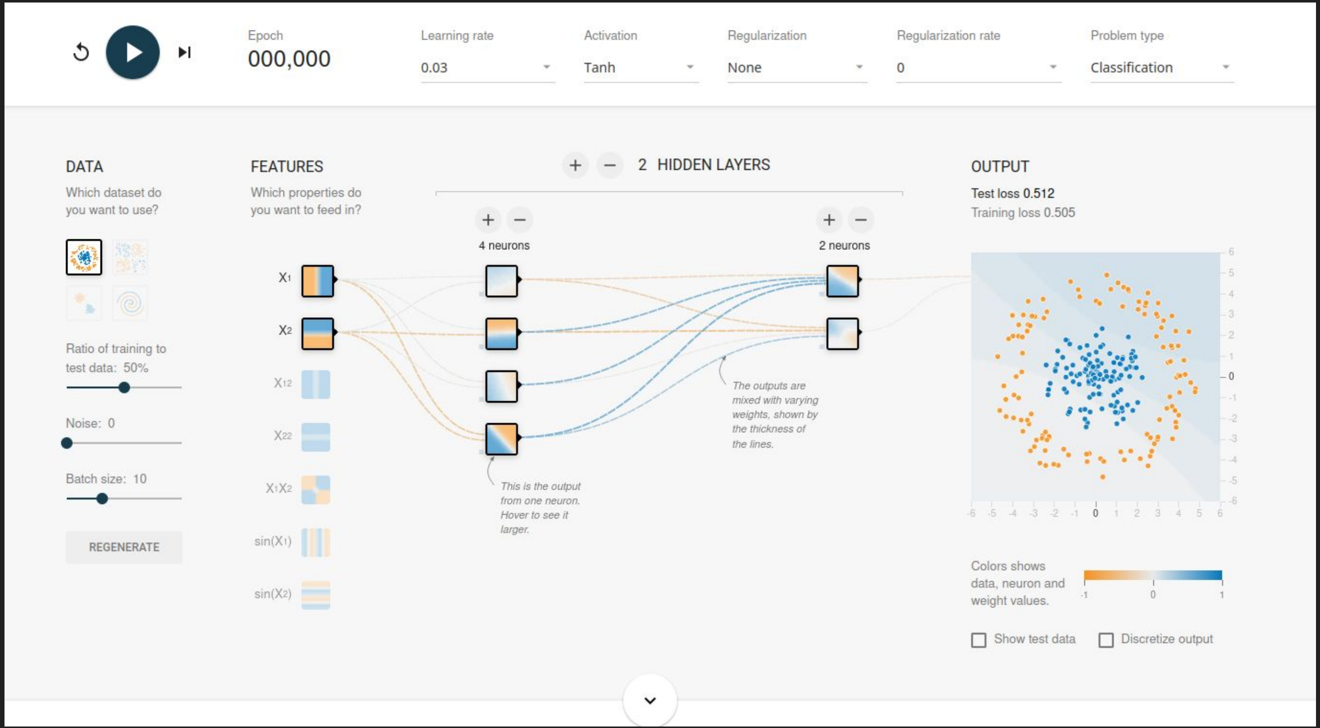
Cat, Dog, Pig, or Plane?

$$p(y0) + p(y1) + p(y2) + p(y3) = 1$$

Classification Demo



Tensorflow Playground



https://playground.tensorflow.org

API

- Datasets:
  https://pytorch.org/vision/stable/datasets.html
- Models:
  https://pytorch.org/vision/stable/models.html
- Tutorials
  https://pytorch.org/tutorials/
- Documentation:
  https://pytorch.org/docs/stable/index.html

HW

https://scikit-learn.org/stable/modules/classes.html#samples-generator

- make_blobs()
- make_circles()
- make_moon()