CMPS 460 – Spring 2022

MACHINE LEARNING

Tamer Elsayed

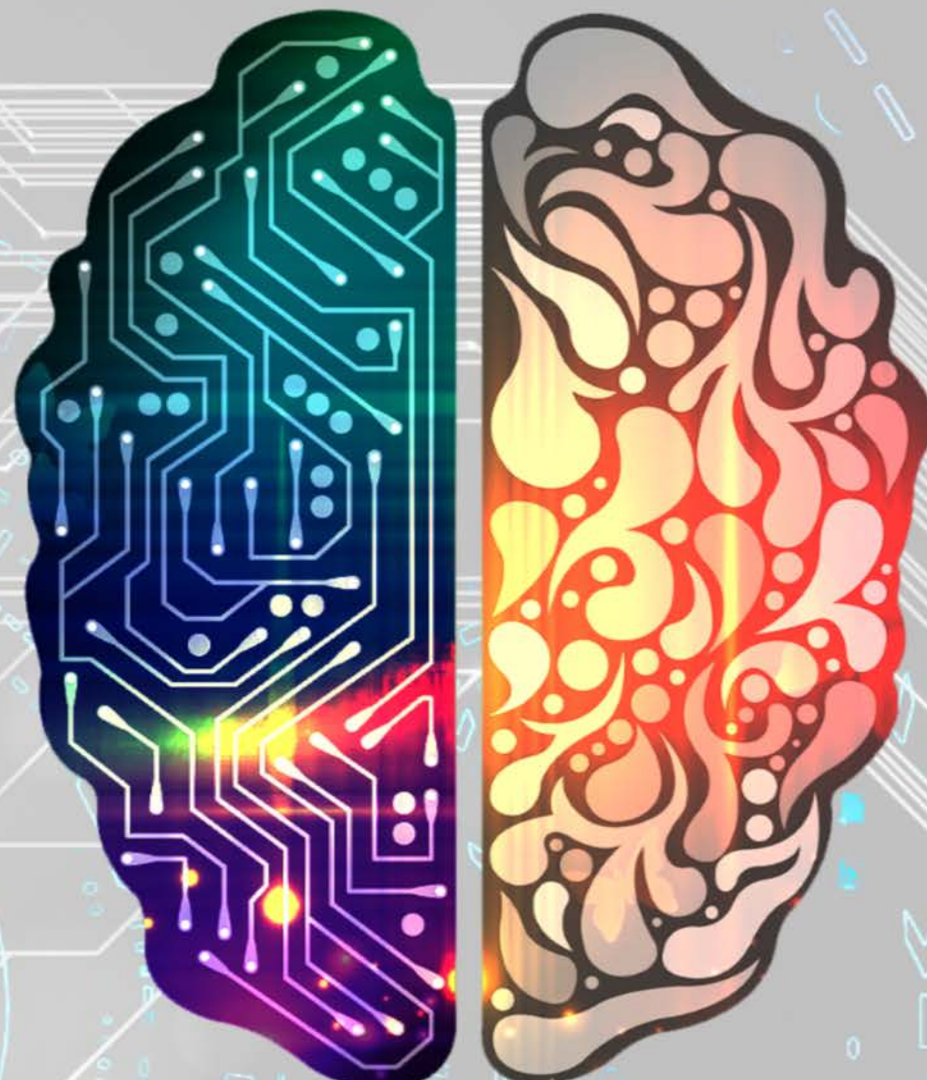


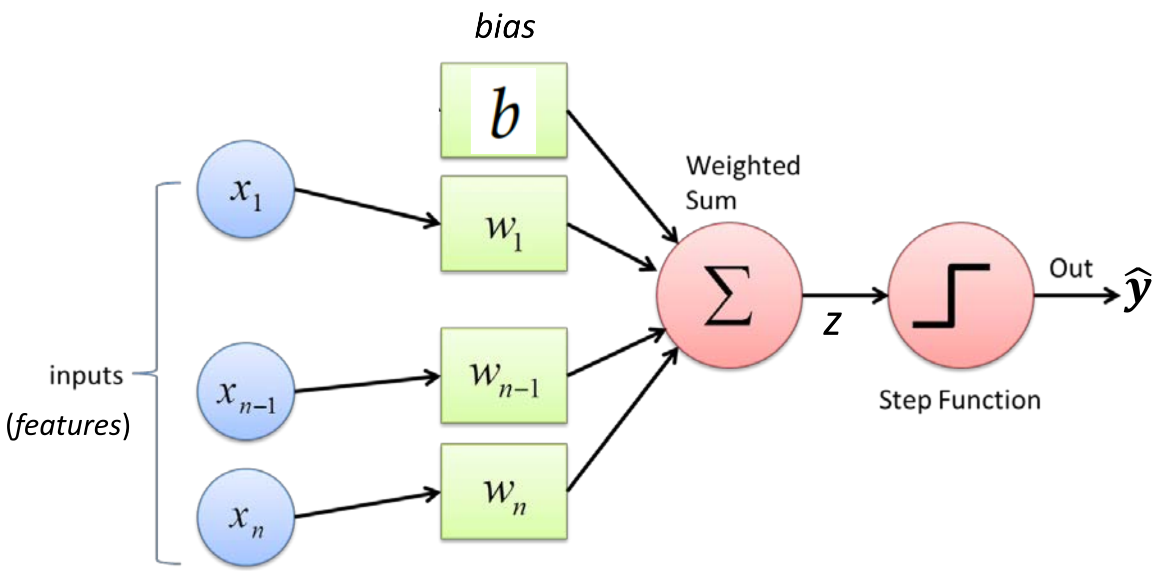
Image hosted by: WittySparks.com | Image source: Pixabay.com

8.c

Prob. Modeling: Logistic Regression

Handout: Intro, 5.1, 5.2.1, 5.4, 5.5, 5.6.1

Perceptron

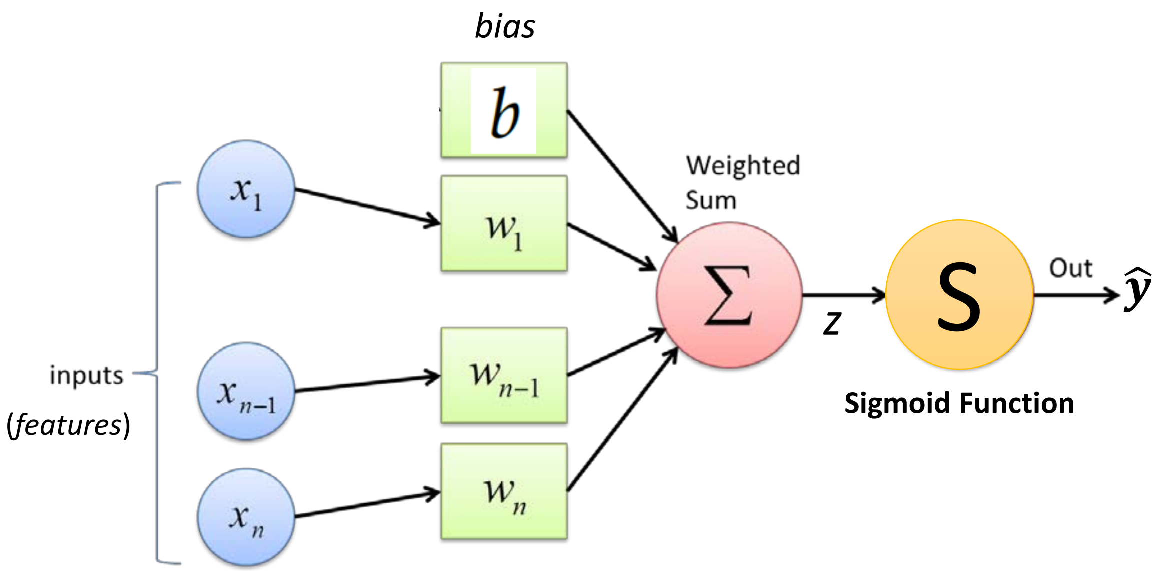


$$z = \left[\sum_{i=1}^{n} w_i x_i\right] + b$$

$$\hat{y} = \mathbb{I}(z > 0)$$

Probability?

Logistic Regression



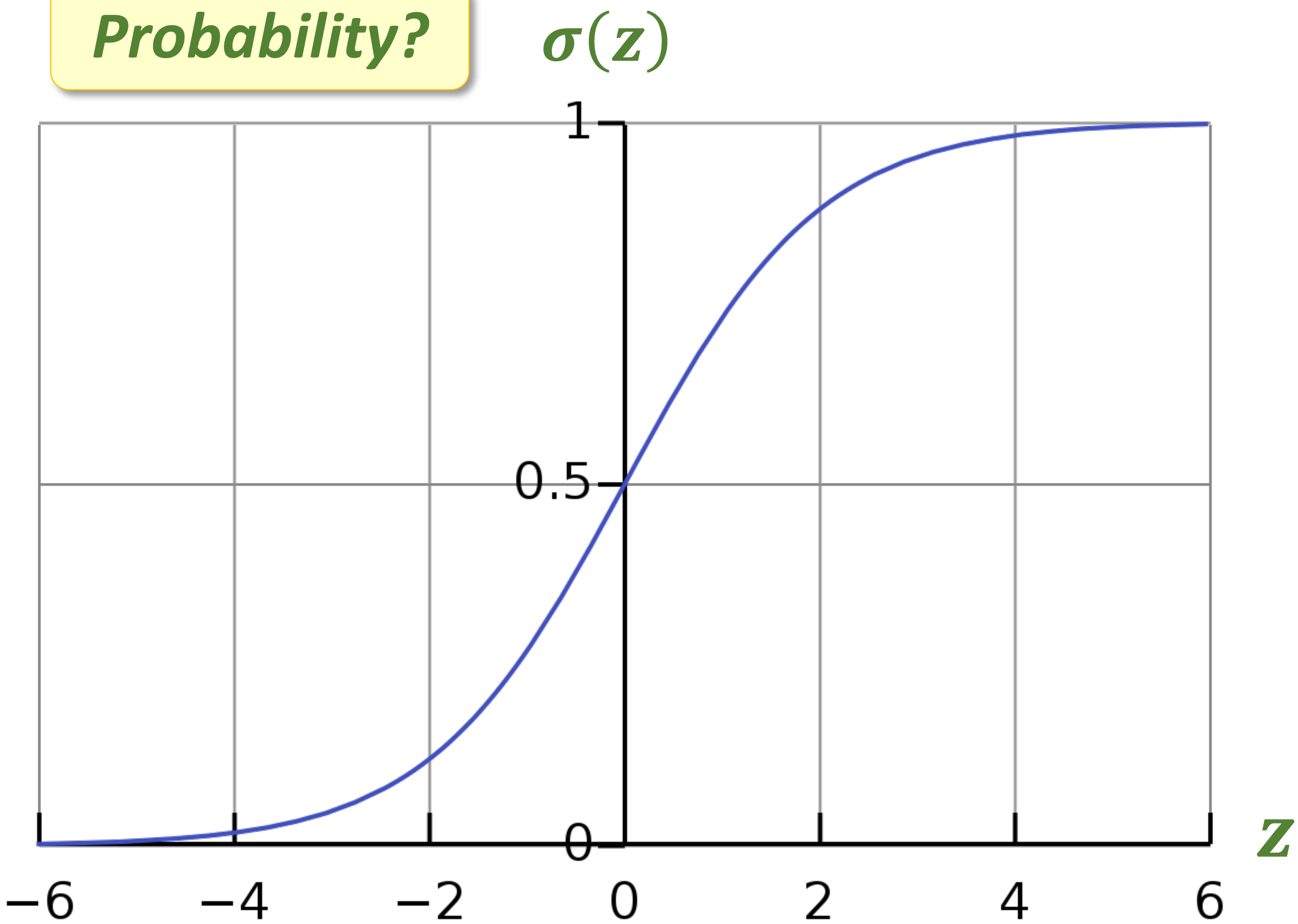
$$z = \left[\sum_{i=1}^{n} w_i x_i \right] + b$$

$$\hat{y} = \sigma(z)$$

Sigmoid Function

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

Probability?



LR Prediction

$$\hat{y} = p(y = 1|x) = \sigma(z) = \sigma(wx + b) = \frac{1}{1 + e^{-(wx+b)}}$$

$$1 - \hat{y} = p(y = 0|x) = 1 - \sigma(z) = 1 - \sigma(wx + b) = \frac{e^{-(wx+b)}}{1 + e^{-(wx+b)}}$$

BTW: $1 - \sigma(z)$=$\sigma(-z)$

$$Predicted\ Label(x) = \begin{cases} 1 & \text{if } \hat{y} > 0.5 \\ 0 & otherwise \end{cases}$$

Example

$n = 3$
$x_1 = 1, x_2 = -1, x_3 = 3$
$w_1 = -1, w_2 = 2, w_3 = 3, b = 0$
$z = ?\ \hat{y} = ?$

if $b = -7$?

Learning

Choosing Parameters …

- Goal: Maximize probability of the correct label
- Remember that

$$\hat{y} = p(y = 1|x)$$

➔ Goal: Maximize $\hat{y}$ when $y = 1$, and $1 - \hat{y}$ when $y = 0$

➔ maximize the following function:

$$p(y|x) \;=\; \hat{y}^{y}\,(1-\hat{y})^{1-y}$$

noting:

if $y = 1$, this simplifies to $\hat{y}$

if $y = 0$, this simplifies to $1 - \hat{y}$

Choose w and b that maximize the probability of the correct labels in the training data

Cross-Entropy Loss

- **Goal**: maximize probability of the correct label $p(y|x)$
- Maximize: $$p(y|x) = \hat{y}^{y}(1-\hat{y})^{1-y}$$
- Maximize: $$\begin{aligned}\log p(y|x) &= \log\left[\hat{y}^{y}(1-\hat{y})^{1-y}\right] \\ &= y\log\hat{y} + (1-y)\log(1-\hat{y})\end{aligned}$$
- Minimize:

$$L_{\mathrm{CE}}(\hat{y},y) = -\log p(y|x) = -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$$

Cross-Entropy loss

Parameters?

- Minimize:

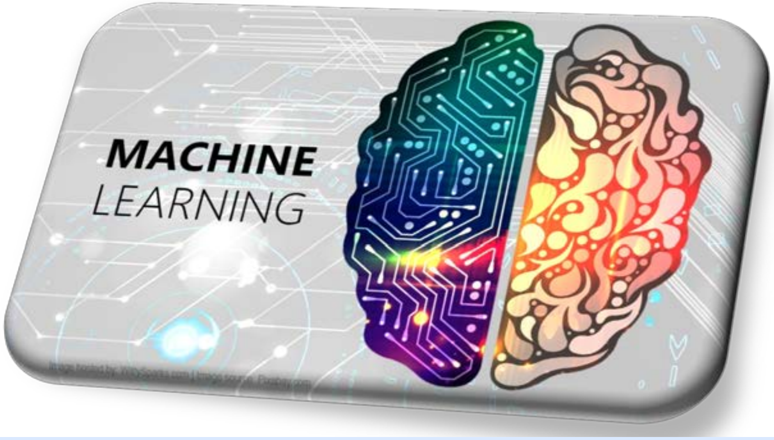
$$L_{\mathrm{CE}}(\hat{y},y) = -\log p(y|x) \;=\; -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$$

- Minimize: ***How?***

$$L_{\mathrm{CE}}(\hat{y},y) \;=\; -[y\log\sigma(w\cdot x+b) + (1-y)\log(1-\sigma(w\cdot x+b))]$$

By Gradient Descent

- Convex and has a unique global minimum
- No closed formula, but solved by gradient descent.



Generative vs. Discriminative Models

Example!

Suppose we're distinguishing cat from dog images

imagenet

imagenet

taken from Jurafsky's http://web.stanford.edu/class/cs124/

Generative Classifier

- Build a model of what's in a cat image
  - Knows about whiskers, ears, eyes
  - Assigns a probability to any image:
    - how cat-y is this image?

- Also build a model for dog images

- Now given a new image:

Run both models and see which one fits better

taken from Jurafsky's http://web.stanford.edu/class/cs124/

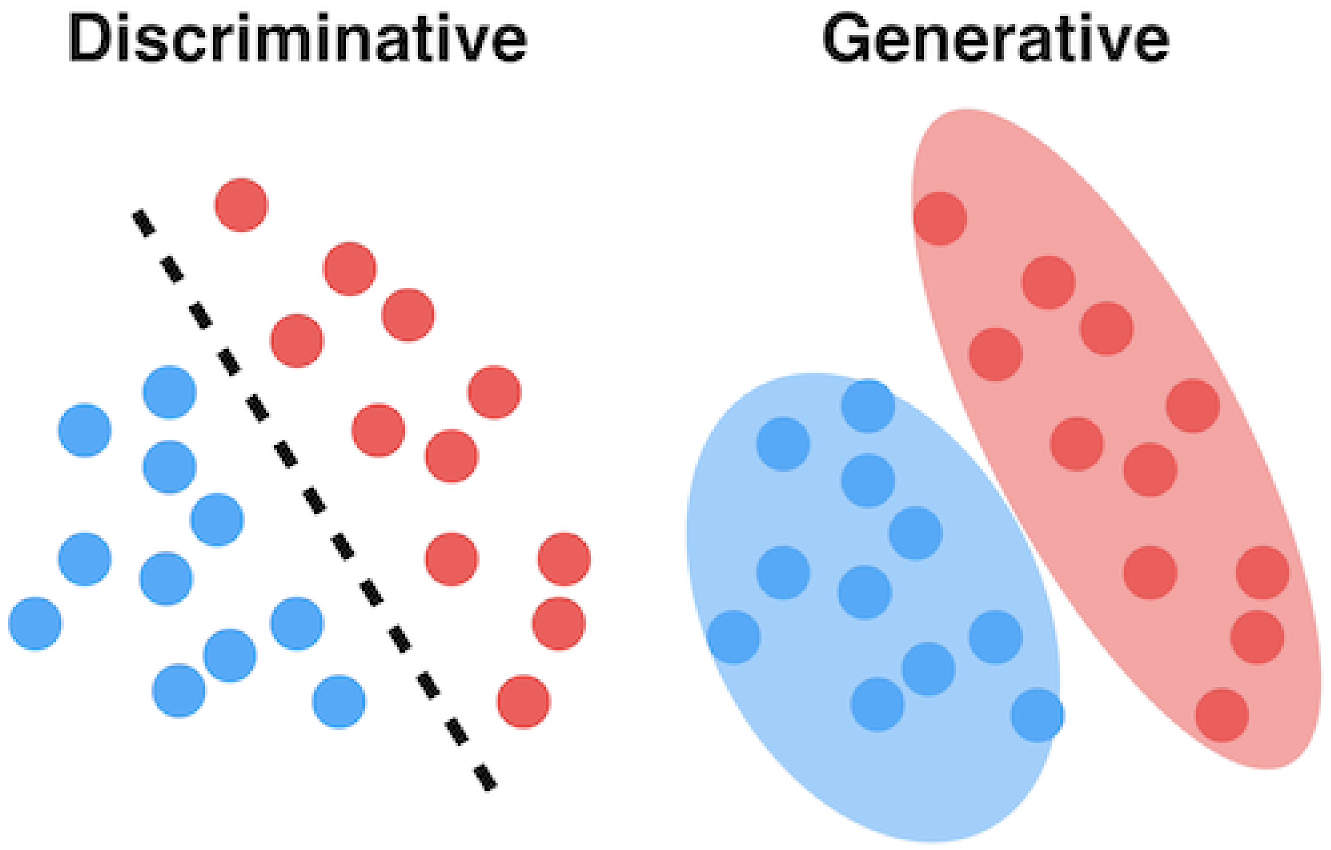
Discriminative Classifier

- Just try to distinguish dogs from cats

Oh look, dogs have collars!
Let's ignore everything else

taken from Jurafsky's http://web.stanford.edu/class/cs124/

Generative vs Discriminative Classifier



https://www.analyticsvidhya.com/blog/2021/07/deep-understanding-of-discriminative-and-generative-models-in-machine-learning/

Example Classifiers

- Naïve Bayes

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} \; \overbrace{P(x|y)}^{\textbf{\textit{likelihood}}} \quad \overbrace{P(y)}^{\textbf{\textit{prior}}}$$

- Logistic Regression

$$\hat{c} = \underset{c \in C}{\operatorname{argmax}} \; \overset{\textbf{\textit{posterior}}}{P(y|x)}$$