CS 260: Foundations of Data Science

Prof. Thao Nguyen

Fall 2026



HAVERFORD
COLLEGE

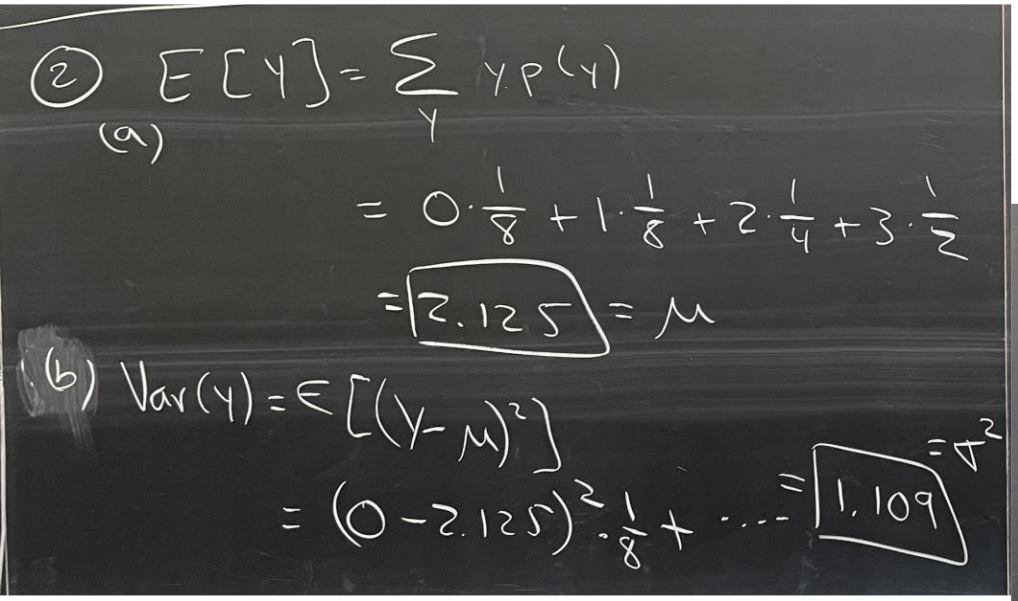
Materials by Sara Mathieson

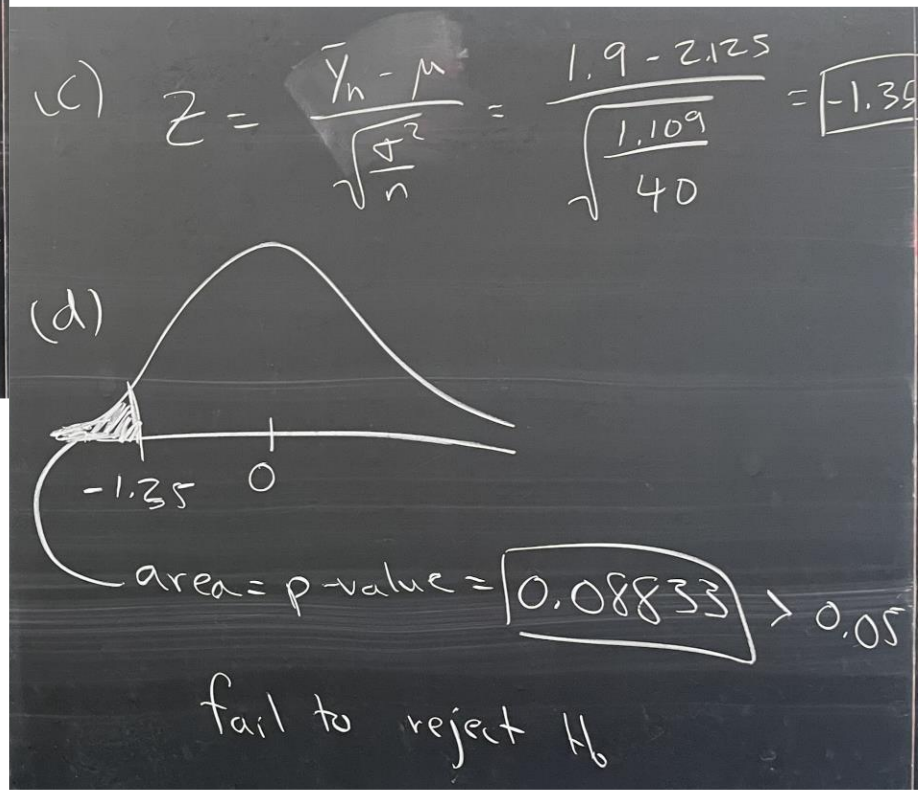
Admin

- **Lab 7** grades & feedback posted on Moodle

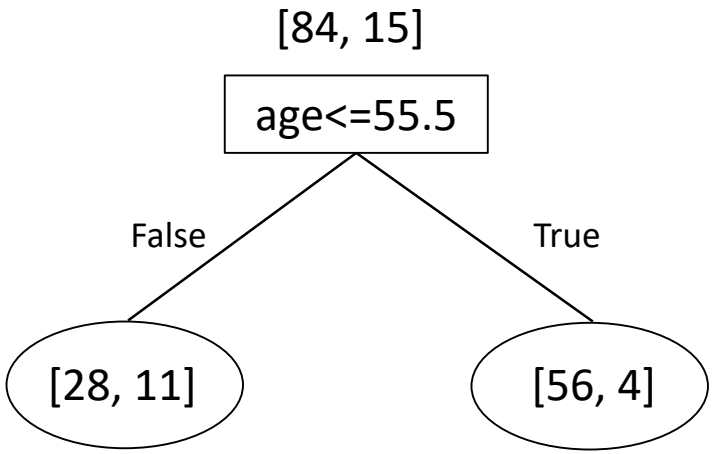
Practice Midterm 2 Solutions

Central Limit Theorem



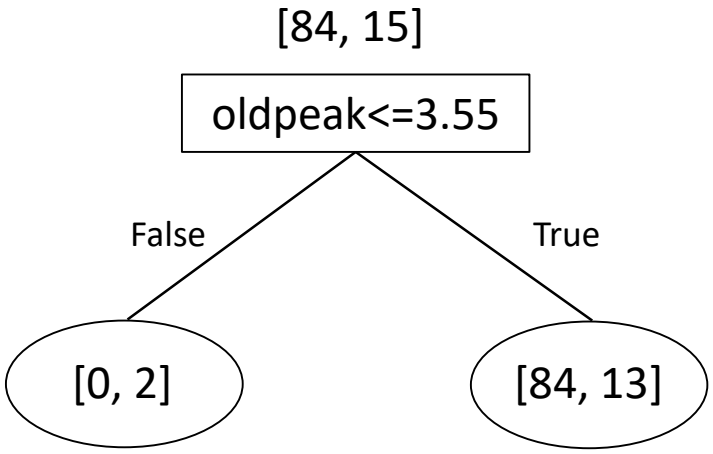


Classification error



Classification error:

$$\frac{11+4}{99} = \frac{15}{99}$$



Classification error: $13/99$

Entropy

- $H(y) = -\sum_{c \in vals(y)} p(y = c) \log_2 \big(p(y = c)\big)$

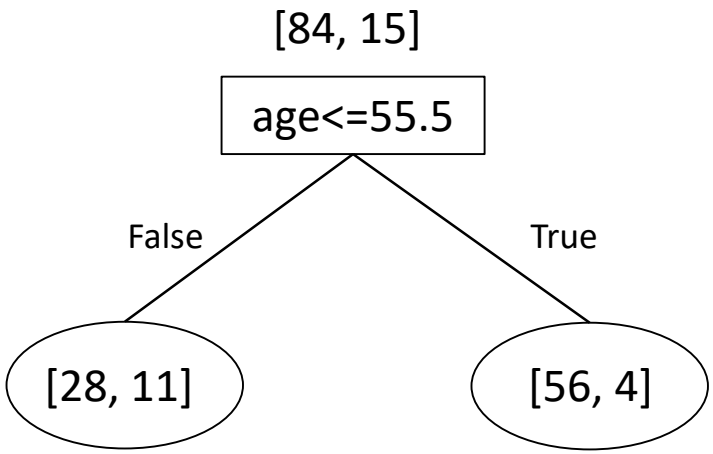
$$= -\left(\frac{84}{99} \log_2 \frac{84}{99} + \frac{15}{99} \log_2 \frac{15}{99}\right) = 0.61$$

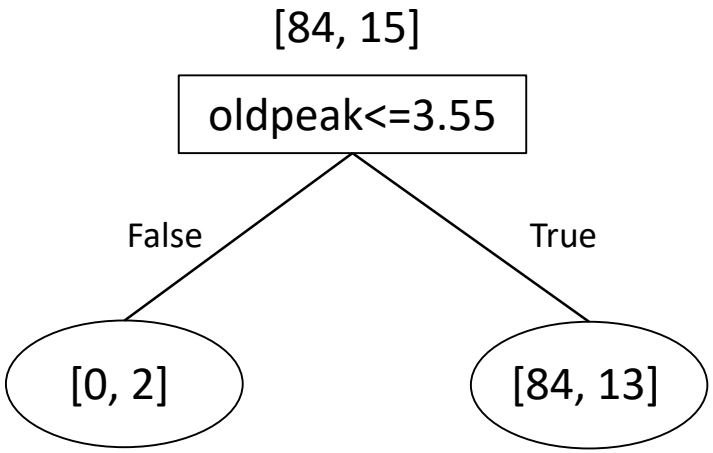
- $H(y|\text{oldpeak}) = \frac{2}{99} H(y|\text{oldpeak} = F) +$

$$\frac{97}{99} H(y|\text{oldpeak} = T)$$

- $H(y|\text{oldpeak} = T) = -\left(\frac{84}{97} \log_2 \frac{84}{97} + \frac{13}{97} \log_2 \frac{13}{97}\right)$

Entropy





H(Y) = 0.613619019599370 8

H(Y|age) = 0.552248091053432 2

H(Y|oldpeak) = 0.556880463059609 3

=> Age feature produces more information gain!

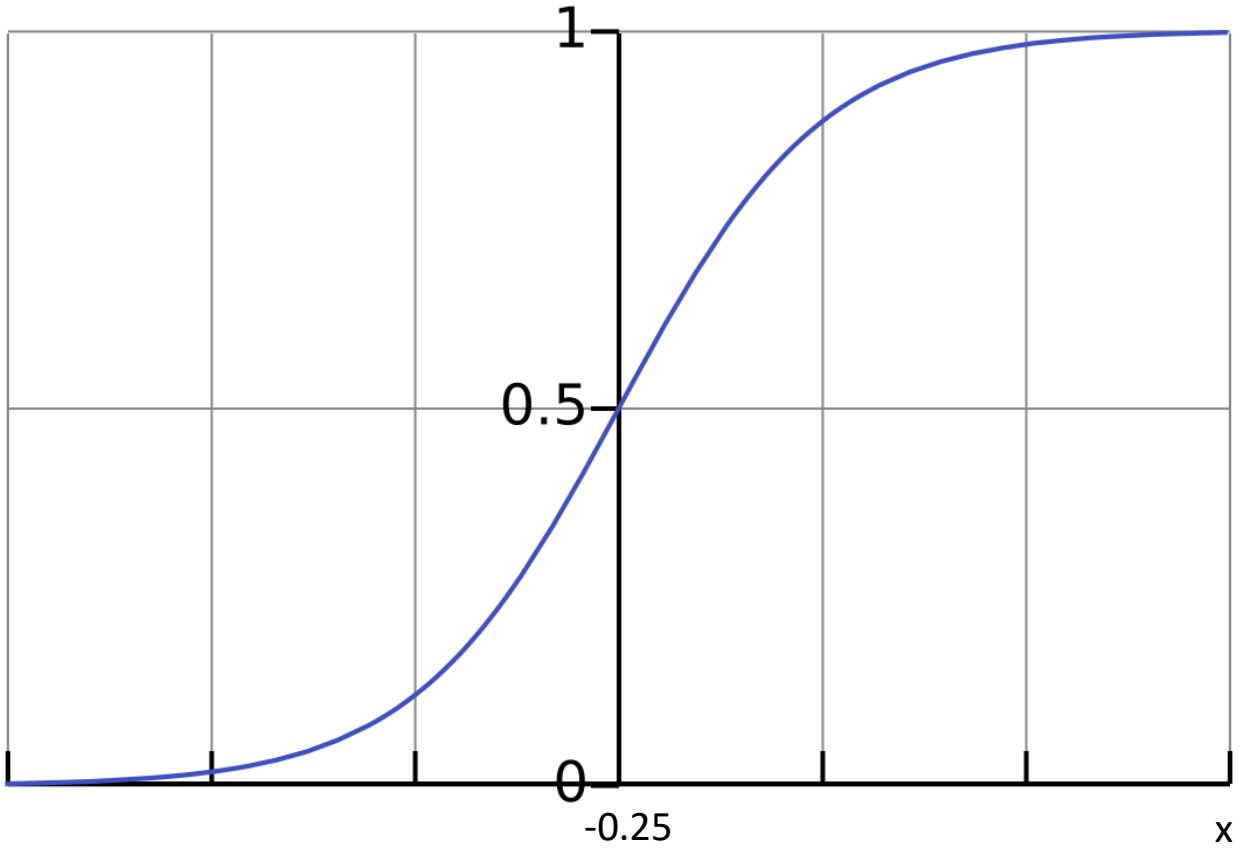
Logistic regression

Say I train a binary logistic regression model (i.e. outcomes $\in \{0, 1\}$) and end up with $\hat{w} = [\hat{w}_0, \hat{w}_1]^T = [1, 4]^T$. What is the decision boundary? Sketch a graph of this logistic model and label the decision boundary. How would you classify a new point $x_{\text{test}} = -0.3$? $< -0.25 \Rightarrow$ predict 0

$$w_0 + w_1 x \geq 0 \Rightarrow \hat{y} = 1$$

$$1 + 4x \geq 0$$

$$x \geq -1/4$$



SGD

We are performing SGD to train a logistic regression model. We start with $\vec{w} = [w_0, w_1]^T = [0, 0]^T$ and $\alpha = 0.01$. What are the new weights after analyzing data point $(x, y) = (-3, 1)$?

$$\vec{w} \leftarrow \vec{w} - \alpha(h_{\vec{w}}(\vec{x_i}) - y_i)\vec{x_i}$$

$$h_{\vec{w}}(\vec{x_i}) = \frac{1}{1 + e^{-\vec{w} \cdot \vec{x_i}}} = \frac{1}{1 + e^0} = 0.5$$

$$\begin{bmatrix} w_0 \\ w_1 \end{bmatrix} \leftarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.01(0.5 - 1)\begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.005 \\ -0.015 \end{bmatrix}$$

Naïve Bayes

$$p(y = k|\boldsymbol{x}) \propto p(y = k) \prod_{j=1}^{p} p(x_j|y = k).$$

- $\theta_0 = \frac{N_0+1}{n+K} = \frac{4}{7}$; $\theta_1 = \frac{3}{7}$

Naïve Bayes

$$p(y = k|\boldsymbol{x}) \propto p(y = k) \prod_{j=1}^{p} p(x_j|y = k).$$

- $\theta_0 = \frac{N_0+1}{n+K} = \frac{4}{7}$; $\theta_1 = \frac{3}{7}$ $\quad \theta_{k,j,v} = \frac{N_{k,j,v} + 1}{N_k + |f_j|}$
- $\vec{x} = [1, D]$
  - $p(\vec{x}|y = 0) = \theta_{0,1,1}\theta_{0,2,D}$

Naïve Bayes

$$p(y = k|\boldsymbol{x}) \propto p(y = k) \prod_{j=1}^{p} p(x_j|y = k).$$

- $\theta_0 = \frac{N_0+1}{n+K} = \frac{4}{7}$; $\theta_1 = \frac{3}{7}$ $\quad \theta_{k,j,v} = \frac{N_{k,j,v} + 1}{N_k + |f_j|}$
- $\vec{x} = [1, D]$
  - $p(\vec{x}|y = 0) = \theta_{0,1,1}\theta_{0,2,D} = \frac{2}{6} * \frac{1}{8} = \frac{1}{24}$
  - $p(\vec{x}|y = 1) = \frac{2}{5} * \frac{1}{7} = \frac{2}{35}$

Naïve Bayes

- $p(y = 0|\vec{x}) \propto \frac{4}{7} * \frac{1}{24} \approx 0.0238$
- $p(y = 1|\vec{x}) \propto \frac{3}{7} * \frac{2}{35} \approx 0.0245$

$\Rightarrow$ predict $y = 1$

Naïve Bayes runtime

What is the runtime of Naive Bayes, assuming each feature has a small, constant number of possible values and the outcome has K possibilities? Also assume that we use **dictionaries** throughout the implementation. Answer for the two steps below in terms of n, p, and K:

i. Estimating the prior probabilities (i.e. computing θ_k for $k = 1, \cdots, K$).

$$\theta_k = \frac{N_k + 1}{n + K} \Rightarrow O(n)$$

ii. Estimating the likelihood probabilities (i.e. computing $\theta_{k,j,v}$ for all k, j, v).

$$\theta_{k,j,v} = \frac{N_{k,j,v} + 1}{N_k + |f_j|} \Rightarrow O(np)$$

Disparate impact

Hypothetically, of the applicants for loans at a bank, 27.5% of the Black applicants got a loan compared to 35% for white applicants. Is there disparate impact in the bank's decisions? Explain your reasoning.

If $P(C = 1|X = 0) < 0.8 * P(C = 1|X = 1)$

$\Rightarrow$ disparate impact

27.5% < 0.8*35%