Fo DA L4
Bayes’ Rule
MLEs and Log-likelihoods
Student Names and Personal Pronouns

Class rosters are provided to the instructor with the students legal name as well as Preferred first name (if previously entered by you in the Student Profile section of your CIS account, which managed can be managed at any time). While CIS refers to this as merely a preference, I will honor you by referring to you with the name and pronoun that feels best for you in class or on assignments. Please advise me of any name or pronoun changes so I can help create a learning environment in which you, your name, and your pronoun are respected. If you need any assistance or support, please reach out to the LGBT Resource Center. https://lgbt.utah.edu/campus/faculty_resources.php

my pronouns: he/him/his

Prof Phillips
\[
\text{Review}
\]

\[
\Pr(\text{P=blue} \land S=\text{green})
\]

\[
\begin{array}{ccc}
\text{S: green} & \text{S: red} & \text{S: blue} \\
\text{P= blue} & 0.3 & 0.1 & 0.2 \\
\text{P= white} & 0.05 & 0.2 & 0.15 \\
\end{array}
\]

\[
\Pr(\text{P=blue} \mid S=\text{red}) = \frac{0.1}{0.3} = 0.3
\]

\[
\Pr(S=\text{red}) = 0.3
\]

\[
\Pr(S=\text{red}) = 0.3
\]

\[
\Pr(S=\text{red}) = 0.3
\]
If $f_X(w) \, dw = 1 \quad w \geq 2 \quad f_X(w) \in [0, \infty)$

R.V.s $X, Y$

$f_{X,Y}(x, y) \rightarrow f_X(x) = \int_{y \geq \gamma} f(x, y) \, dy$

$f_{X|Y}(x, y=y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$

$f_Y(y) = \int_{x \in \mathbb{R}} f(x, y) \, dx$
Bayes' Rule

\[ P_r(M | D) = \frac{P_r(D | M) \cdot P_r(M)}{P_r(D)} \]

Two events M, D

\[ P_r(M | D) = \frac{P_r(M \cap D)}{P_r(D)} \]

\[ P_r(M \cap D) = P_r(M | D) \cdot P_r(D) \]

\[ P_r(M | D) = P_r(D | M \cap D) = P_r(D | M) \cdot P_r(M) \]

\[ P_r(M | D) \cdot P_r(D) = P_r(D | M) \cdot P_r(M) \]
\[
\Pr(C | D) = \frac{\Pr(C | D \cap M) \cdot \Pr(M)}{\Pr(C | D \cap M) \cdot \Pr(M) + \Pr(C | D \cap \overline{M}) \cdot \Pr(\overline{M})}
\]

\[
\Pr(C | D) = \frac{\Pr(C | D \cap M) \cdot \Pr(M)}{\Pr(C | D \cap M) \cdot \Pr(M) + \Pr(C | D \cap \overline{M}) \cdot \Pr(\overline{M})}
\]

\[
\Pr(C | D) = \frac{\Pr(C | D \cap M) \cdot \Pr(M)}{\Pr(C | D \cap M) \cdot \Pr(M) + \Pr(C | D \cap \overline{M}) \cdot \Pr(\overline{M})}
\]

\[
\Pr(C | D) = \frac{\Pr(C | D \cap M) \cdot \Pr(M)}{\Pr(C | D \cap M) \cdot \Pr(M) + \Pr(C | D \cap \overline{M}) \cdot \Pr(\overline{M})}
\]

\[
\Pr(C | D) = \frac{\Pr(C | D \cap M) \cdot \Pr(M)}{\Pr(C | D \cap M) \cdot \Pr(M) + \Pr(C | D \cap \overline{M}) \cdot \Pr(\overline{M})}
\]

\[
\Pr(C | D) = \frac{\Pr(C | D \cap M) \cdot \Pr(M)}{\Pr(C | D \cap M) \cdot \Pr(M) + \Pr(C | D \cap \overline{M}) \cdot \Pr(\overline{M})}
\]
Cracked windshield

Event: W windshield cracked.

3 factors: A, B, C

\[ \Pr(A) = 0.5 \quad \Pr(B) = 0.3 \quad \Pr(C) = 0.2 \]

\[ \Pr(w|A) = 0.01 \quad \Pr(w|B) = 0.1 \quad \Pr(w|C) = 0.02 \]

\[
\Pr(A|w) = \frac{\Pr(w|A) \cdot \Pr(A)}{\Pr(w)} = \frac{(0.01)(0.5)}{\Pr(w)} = \frac{0.005}{\Pr(w)}
\]

\[
\Pr(B|w) = \frac{\Pr(w|B) \cdot \Pr(B)}{\Pr(w)} = \frac{(0.1)(0.3)}{\Pr(w)} = \frac{0.03}{\Pr(w)}
\]

\[
\Pr(C|w) = \frac{\Pr(w|C) \cdot \Pr(C)}{\Pr(w)} = \frac{(0.02)(0.2)}{\Pr(w)} = \frac{0.004}{\Pr(w)}
\]
$M = \text{model}$

$D = \text{data}$

$P_r(M|D)$

$M \in \mathcal{M}$

specify the model

$L_{\text{MAP}}$ maximizes $P_r(M|D)$

$L_{M^*}$ MAP maximum a posteriori

$M^* = \arg \max_{M \in \mathcal{M}} P_r(M|D)$

$P_r(M) = \frac{P_r(D|M) \cdot P_r(M)}{P_r(D)}$

$\text{MLE}$ assume

$M^* = \arg \max_{M \in \mathcal{M}} P_r(M)$

$P_r(M|D) = \arg \max_{M \in \mathcal{M}} P_r(D|M) \\ P_r(M)$

maximum likelihood

posterior likelihood
\[ D = \{ x_1, x_2, \ldots, x_n \} \] independent observations

\( M \) - explaining structure in data

1. DCR': Model: \( M \) = value of average height (Normally distributed)

2. Linear regression

\[ M \text{ line in } \mathbb{R}^2 \]

3. Clustering

\( \text{D } \in \mathbb{R}^d \)

\( M \) set of points
4. **PCA**  \( D \in \mathbb{R}^d \)  \( M \in F \) \( k \)-dimensional subspace

5. (linear) **classification**  

\( D \in \mathbb{R}^{d \times 3-1, +13} \)  
\( x_i \in \mathbb{R}^d \)  
\( y \in \{3, -1, +1\} \)
\[ D \subset \mathbb{R}^n = \{x_1, x_2, \ldots, x_n\} = \{1, 3, 12, 5, 9\} \]

**Modeling**

\[ x_i \sim N(\mu, \sigma) \]

**Model** \( M = m \)

\[
m \in \mathbb{R} = \mathcal{M}
\]

\[
N_{m, \sigma^2}(x) = g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(m-x)^2\right)
\]

"Pr" \( x_i = x \mid M = m \)

\[
\Pr(D|M) = \prod_{x_i \in D} g(x_i) = \prod_{x_i \in D} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(m-x_i)^2\right)\right)
\]
\[ Pr(D|M) = \prod_{x \in D} g(x_i) = \prod_{x \in D} \left( \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} (m-x_i)^2 \right) \right) \]

\[ \arg \max_{m \in \mathbb{R}} m \]

\[ Pr(D|M) = \arg \max_{m \in \mathbb{R}} \log \left( Pr(D|M) \right) \]

\[ \log(a \cdot b) = \log(a) + \log(b) \]

\[ = \arg \max_{m} \log \left( \prod_{x_i} g(x_i) \right) = \arg \max_{m} \frac{1}{m} \sum_{x_i} \log(g(x_i)) \]

\[ = \arg \max_{m} \frac{1}{m} \sum_{x_i} \left( \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} (m-x_i)^2 \right) \right) \]

\[ = \text{average} \left( x_1, \ldots, x_n \right) \]