Reminders:

Homework 5 is out

Quiz 5 Question 3 getting regraded
Stay tuned – figuring out the Canvas side of regrading
Supervised Learning: Naïve Bayes

General probabilistic approach:

\[
\arg\max\limits_{c\in C} P(c|\bar{x}_{\text{test},1}, \ldots, \bar{x}_{\text{test},n}) \\
= \arg\max\limits_{c\in C} P(c) P(\bar{x}_{\text{test},1}, \ldots, \bar{x}_{\text{test},n}|c)
\]
Supervised Learning: Naïve Bayes

General probabilistic approach:

$$\arg\max_{c \in C} P(c | \bar{x}_{test,1}, \ldots, \bar{x}_{test,n})$$

$$= \arg\max_{c \in C} P(c) P(\bar{x}_{test,1}, \ldots, \bar{x}_{test,n} | c)$$

Naïve Bayes: $$\arg\max_{c \in C} P(c) \prod_{j=1}^{n} P(\bar{x}_{test,j} | c)$$
Supervised Learning: Naïve Bayes

General probabilistic approach:

\[
\arg\max_{c \in C} P(c | \tilde{x}_{test,1}, ..., \tilde{x}_{test,n}) \\
= \arg\max_{c \in C} P(c) P(\tilde{x}_{test,1}, ..., \tilde{x}_{test,n} | c)
\]

Naïve Bayes: \(\arg\max_{c \in C} P(c) \prod_{j=1}^{n} P(\tilde{x}_{test,j} | c)\)

Estimate using training data: \(\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_lj}}{N_l} \right]\)
Supervised Learning: Naïve Bayes

Estimate using training data:

$$\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_jl}}{N_l} \right]$$
Supervised Learning: Naïve Bayes

Estimate using training data:

$$\arg \max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_l}}{N_l} \right]$$

Avoiding underflow:

$$\arg \max_{c_l \in C} \left[ \left( \sum_{j=1}^{n} \log n_{ja_l} \right) - (n - 1) \log N_l \right]$$
Supervised Learning: Naïve Bayes

Estimate using training data:

$$\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_l}}{N_l} \right]$$

Binary-valued attributes:

$$\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \left( \frac{n_{j1l}}{N_l} \right)^{x_j} \left( 1 - \frac{n_{j1l}}{N_l} \right)^{1-x_j} \right]$$
Supervised Learning: Naïve Bayes

$$\underset{c_l \in C}{\text{argmax}} \left( \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_jl}}{N_l} \right)$$
Supervised Learning: Naïve Bayes

\[
\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_j l}}{N_l} \right]
\]

Vector space model of text: \( V_j = \{0,1\} \) presence or absence of word j
Supervised Learning: Naïve Bayes

\[ \arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_l}}{N_l} \right] \]

Vector space model of text: \( V_j = \{0,1\} \)  presence or absence of word \( j \)

What happens if for some category \( c \) there are no occurrences of word \( j \)?
Supervised Learning: Naïve Bayes

\[
\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_l}}{N_l} \right]
\]

Vector space model of text: \( V_j = \{0,1\} \) presence or absence of word j

What happens if for some category \( c_l \) there are no occurrences of word j? 
\( n_{ja_l} = 0 \)
Supervised Learning: Naïve Bayes

\[
\arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_i l}}{N_l} \right]
\]

If any of these is 0 then the whole thing is 0

Vector space model of text: \( V_j = \{0,1\} \) presence or absence of word \( j \)

What happens if for some category \( c_l \) there are no occurrences of word \( j \)?

\[ n_{ja_i l} = 0 \]
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Intuition: Let’s start out as if all values are equally likely \( \frac{1}{|V_j|} \)
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Intuition: Let’s start out as if all values are equally likely ($\frac{1}{|V_j|}$)

Before: Estimate $P(\tilde{x}_{test,j} = v_a | c_l) = \frac{n_{jal}}{N_l}$

What if we use: $P(\tilde{x}_{test,j} = v_a | c_l) = \frac{n_{jal}+1}{N_l+|V_j|}$?
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Intuition: Let’s start out as if all values are equally likely ($\frac{1}{|V_j|}$)

Before: Estimate $P(\tilde{x}_{test,j} = v_a \mid c_l) = \frac{n_{jal}}{N_l}$

What if we use: $P(\tilde{x}_{test,j} = v_a \mid c_l) = \frac{n_{jal}+1}{N_l+|V_j|}$?

Acts as if we’ve seen one occurrence of each value
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Intuition: Let’s start out as if all values are equally likely ($\frac{1}{|V_j|}$)

Before: Estimate $P(\tilde{x}_{test,j} = v_a | c_l) = \frac{n_{jal}}{N_l}$

What if we use: $P(\tilde{x}_{test,j} = v_a | c_l) = \frac{n_{jal}+1}{N_l+|V_j|}$?

Acts as if we’ve seen one occurrence of each value "pseudocount"
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Estimate $P(\tilde{x}_{test,j} = v_a | c_l)$ using $\frac{n_{jal} + \alpha}{N_l + \alpha |V_j|}$
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Estimate $P(\tilde{x}_{test,j} = v_a | c_l)$ using $\frac{n_{jat} + \alpha}{N_l + \alpha |V_j|}$

$\alpha > 0$: “smoothing parameter”
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Estimate $P(\tilde{x}_{test,j} = v_a | c_l)$ using

$$\frac{n_{jat} + \alpha}{N_l + \alpha |V_j|}$$

$\alpha > 0$: “smoothing parameter”

(Often $\alpha \leq 1$)
Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

Estimate \( P(\tilde{x}_{test,j} = v_a \mid c_l) \) using \( \frac{n_{jat} + \alpha}{N_l + \alpha |V_j|} \)

\( \alpha > 0 \): “smoothing parameter”

(Often \( \alpha \leq 1 \))

(seems like a “trick” but there are probabilistic justifications)
Naïve Bayes: Laplace Smoothing

Attributes from discrete sets: Assign label $c$

$$c = \arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_jl} + \alpha}{N_l + \alpha |V_j|} \right]$$

Binary attributes case: Assign label $c$

$$c = \arg\max_{c_l \in C} \left[ \frac{N_l}{N} \prod_{j=1}^{n} \left( \frac{n_{j1l} + \alpha}{N_l + \alpha |V_j|} \right)^{x_j} \left( 1 - \frac{n_{j1l} + \alpha}{N_l + \alpha |V_j|} \right)^{1-x_j} \right]$$