CS 4700: Foundations of Artificial Intelligence

Spring 2020
Prof. Haym Hirsh

Lecture 33
May 1, 2020

ONLINE

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:

$$\underset{c \in C}{\operatorname{argmax}}\, P\left(c \middle| \bar{x}_{test,1}, \dots, \bar{x}_{test,n}\right)$$
$$= \underset{c \in C}{\operatorname{argmax}}\; P(c)\, P(\bar{x}_{test,1}, \dots, \bar{x}_{test,n} | c)$$

Supervised Learning: Naïve Bayes

General probabilistic approach:

$$\underset{c \in C}{\operatorname{argmax}}\, P\left(c \middle| \bar{x}_{test,1}, \dots, \bar{x}_{test,n}\right)$$
$$= \underset{c \in C}{\operatorname{argmax}}\; P(c)\, P(\bar{x}_{test,1}, \dots, \bar{x}_{test,n} | c)$$

Naïve Bayes: $\underset{c \in C}{\operatorname{argmax}}\, P(c) \prod_{j=1}^{n} P(\bar{x}_{test,j} | c)$

Supervised Learning: Naïve Bayes

General probabilistic approach:

$$\underset{c \in C}{\operatorname{argmax}}\, P\left(c \middle| \bar{x}_{test,1}, \dots, \bar{x}_{test,n}\right)$$
$$= \underset{c \in C}{\operatorname{argmax}}\; P(c)\, P(\bar{x}_{test,1}, \dots, \bar{x}_{test,n} | c)$$

Naïve Bayes: $\underset{c \in C}{\operatorname{argmax}}\, P(c) \prod_{j=1}^{n} P(\bar{x}_{test,j} | c)$

Estimate using training data: $\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_j l}}{N_l} \right]$

Supervised Learning: Naïve Bayes

Estimate using training data:

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_j l}}{N_l} \right]$$

Supervised Learning: Naïve Bayes

Estimate using training data:

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_j l}}{N_l} \right]$$

Avoiding underflow:

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\left(\sum_{j=1}^{n} \log\ n_{ja_j l} \right) - (n-1) \log N_l \right]$$

Supervised Learning: Naïve Bayes

Estimate using training data:

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_j l}}{N_l} \right]$$

Binary-valued attributes:

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{\mathrm{j}=1}^{\mathrm{n}} \left(\frac{n_{j1l}}{N_l} \right)^{\mathrm{x_j}} \left(1 - \frac{n_{j1l}}{N_l} \right)^{1-\mathrm{x}_j} \right]$$

Supervised Learning: Naïve Bayes

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_jl}}{N_l} \right]$$

Supervised Learning: Naïve Bayes

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{j a_j l}}{N_l} \right]$$

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

Supervised Learning: Naïve Bayes

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{j a_j l}}{N_l} \right]$$

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

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{n_{ja_jl}}{N_l} \right]$$

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?

$$n_{ja_jl} = 0$$

Supervised Learning: Naïve Bayes

$$\underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{j=1}^{n} \frac{\boxed{n_{ja_j l}}}{N_l} \right]$$

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?

$$n_{ja_j l} = 0$$

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 $P\left(\bar{x}_{test,j} = v_a \,|c_l\right) = \frac{n_{jal}}{N_l}$

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

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 $P\left(\bar{x}_{test,j} = v_a \,|c_l\right) = \frac{n_{jal}}{N_l}$

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

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 $P\left(\bar{x}_{test,j} = v_a \,|c_l\right) = \frac{n_{jal}}{N_l}$

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

Acts as if we’ve seen one occurrence of each value

"pseudocount"

Supervised Learning: Naïve Bayes

Laplace (or “Additive”) Smoothing

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

Supervised Learning: Naïve Bayes

Laplace (or "Additive") Smoothing

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

$\alpha > 0$: "smoothing parameter"

Supervised Learning: Naïve Bayes

Laplace (or "Additive") Smoothing

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

$\alpha > 0$: "smoothing parameter"

(Often $\alpha \leq 1$)

Supervised Learning: Naïve Bayes

Laplace (or "Additive") Smoothing

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

$\alpha > 0$: "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

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

Binary attributes case: Assign label c

$$c = \underset{c_l \in C}{\operatorname{argmax}} \left[\frac{N_l}{N} \prod_{\mathrm{j}=1}^{\mathrm{n}} \left(\frac{n_{j1l} + \alpha}{N_l + \alpha |V_j|} \right)^{\mathrm{x_j}} \left(1 - \frac{n_{j1l} + \alpha}{N_l + \alpha |V_j|} \right)^{1-\mathrm{x}_j} \right]$$