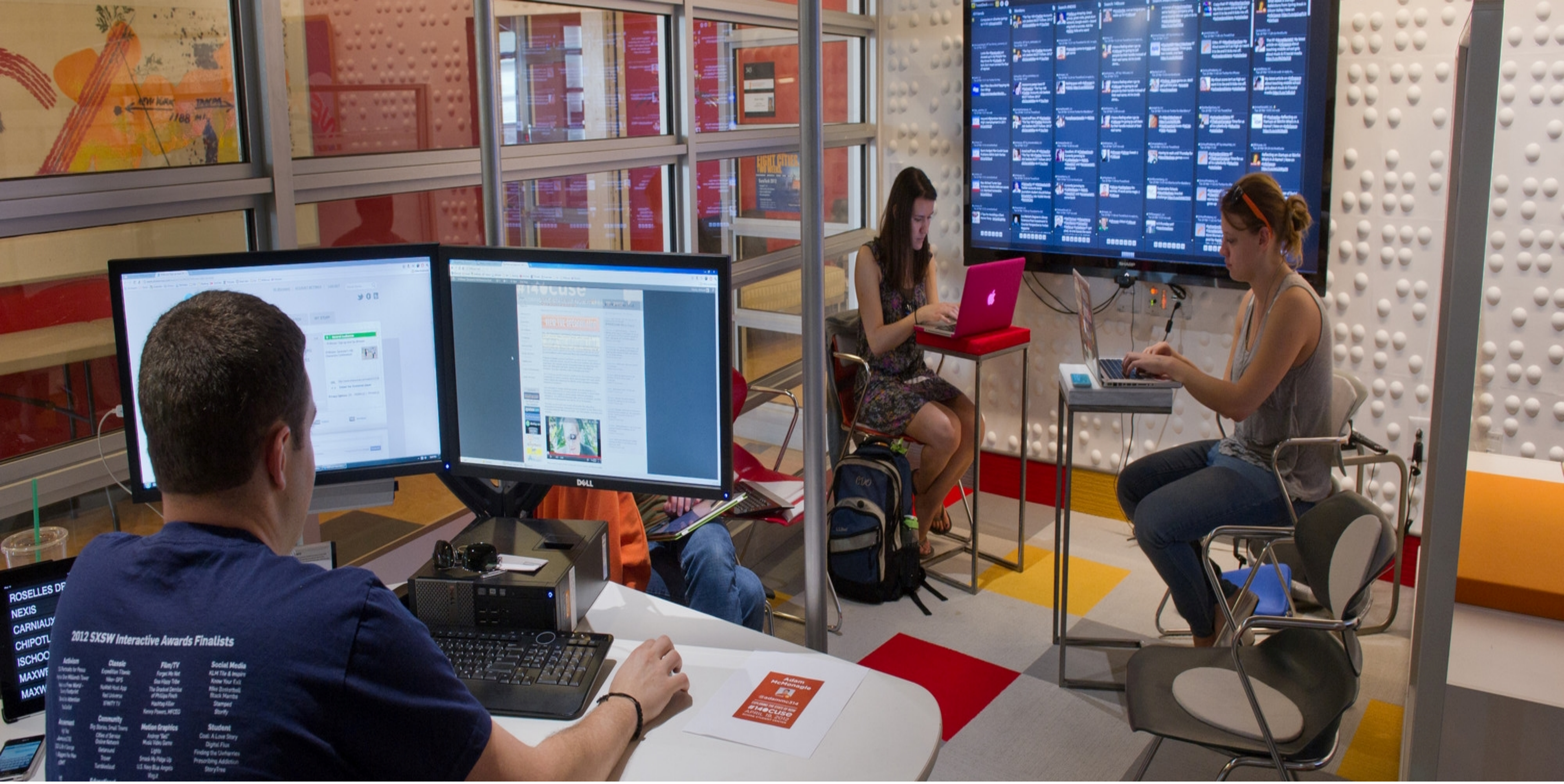

SMOOTHING

SYRACUSE UNIVERSITY
School of Information Studies

PROBLEM OF ZERO PROBABILITIES

$$
\begin{aligned}
&P(cancer = no \mid pos\_mammo = yes, fam\_hist = yes, caucasian = yes) \\
&= \frac{P(pos\_mammo = yes \mid cancer = no) \cdot P(fam\_hist = yes \mid cancer = no) \cdot P(caucasian = yes \mid cancer = no) \cdot P(cancer = no)}{P(pos\_mammo = yes, fam\_hist = yes, caucasian = yes)} \\
&= \frac{(2/4) \cdot (0/4) \cdot (2/4) \cdot (4/7)}{P(pos\_mammo = yes, fam\_hist = yes, caucasian = yes)} \\
&= 0
\end{aligned}
$$

Zero probability of no cancer?

PROBLEM OF ZERO PROBABILITIES

If one of the conditional probabilities is zero, then the entire product becomes zero.

But the zero probability may just be caused by lack of data.

P(fam_hist=yes | cancer=yes) = 1
P(fam_hist=no | cancer=yes) = 0
P(fam_hist=yes | cancer=no) = 0
P(fam_hist=no | cancer=no) = 1

SOLUTION TO ZERO PROBABILITY

Probability estimation should replace these zero probabilities with a very small probability, which means such events still occur in real world but are so rare that the training data did not include any of them.

Since all probabilities should add up to one, the other nonzero probabilities need to “shrink” a little bit, in order to “lend” the small amount to the zero probabilities.

This technique is called “smoothing.”

SMOOTHING FOR ZERO PROBABILITIES

Using a smoothing algorithm, such as Laplacian smoothing, also called "add-one" smoothing

$$\boxed{\begin{aligned} \text{Original}: P(A_i \mid C) &= \frac{N_{ic}}{N_c} \\ \text{Laplace}: P(A_i \mid C) &= \frac{N_{ic}+1}{N_c + c} \end{aligned}}$$

A binary classifier has two classes (c = 2).

ADD-ONE SMOOTHING

Example	Pos_mammo	Fam_hist	Alcohol	Cancer
1	Yes	Yes	Yes	Yes
2	Yes	Yes	No	Yes
3	No	Yes	Yes	Yes
4	Yes	No	No	No
5	Yes	No	Yes	No
6	No	No	Yes	No
7	No	No	No	No

	Cancer = Yes	Cancer = No
Fam_hist = Yes	3	0
Fam_hist = No	0	4

→

	Cancer = Yes	Cancer = No
Fam_hist = Yes	3 + 1	0 + 1
Fam_hist = No	0 + 1	4 + 1

Add-one smoothing means to add an example to each category.

ADD-ONE SMOOTHING

	Cancer = Yes	Cancer = No
Fam_hist = Yes	3	0
Fam_hist = No	0	4

→

	Cancer = Yes	Cancer = No
Fam_hist = Yes	3 + 1	0 + 1
Fam_hist = No	0 + 1	4 + 1

Add-one smoothing means to add an example to each category.

Original probabilities:
P(fam_hist=yes | cancer=yes) = 3/3
P(fam_hist=no | cancer=yes) = 0/3 = 0
P(fam_hist=yes | cancer=no) = 0/4 = 0
P(fam_hist=no | cancer=no) = 4/4 = 1

→

Smoothed probabilities:
P(fam_hist=yes | cancer=yes) = 4/5
P(fam_hist=no | cancer=yes) = 1/5
P(fam_hist=yes | cancer=no) = 1/6
P(fam_hist=no | cancer=no) = 5/6

The nonzero probabilities become smaller, “lending” values to zero probabilities.

SMOOTHING FOR ZERO PROBABILITIES

Using a smoothing algorithm, such as Laplacian smoothing, also called “add-one” smoothing

$$\text{Original}: P(A_i \mid C) = \frac{N_{ic}}{N_c}$$

$$\text{Laplace}: P(A_i \mid C) = \frac{N_{ic} + 1}{N_c + c}$$

A binary classifier has two classes (c = 2).

$$\frac{0}{3} = 0$$

$$\frac{0+1}{3+2} = \frac{1}{5} = 0.2$$

Well, this is not a very small probability!

SMOOTHING FOR ZERO PROBABILITIES

Using a smoothing algorithm, such as Laplacian smoothing, also called “add-one” smoothing

$$\text{Original}: P(A_i \mid C) = \frac{N_{ic}}{N_c}$$

$$\text{Laplace}: P(A_i \mid C) = \frac{N_{ic} + 1}{N_c + c}$$

The probabilities in real-world problems are usually very small; thus, add-one smoothing would change the probability just a little bit.

$$\frac{1}{7000} = 0.000143$$

$$\frac{1+1}{7000+2} = \frac{2}{7002} = 0.000285$$

LOG PROBABILITIES

The probabilities are so small that we usually use log(P) instead, so that we don’t have to store that many preceding zeros, as in 0.000222.

$$\boxed{\begin{aligned} &\log\left(\frac{1}{7000}\right) = \quad 8.854 \\ &\log\left(\frac{1+1}{7000+2}\right) = \log\left(\frac{2}{7002}\right) = \quad 8.161 \end{aligned}}$$