MIDS W207
Applied Machine Learning

Spring 2022

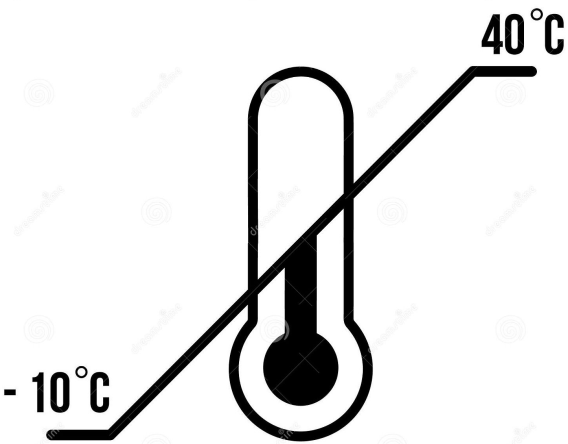
Week 3
Live Session Slides

Probability Review

- Random Variable
- Probability Distribution
- Joint Distribution
- Marginal Distribution

Probability Review

- Random Variable
- Probability Distribution
- Joint Distribution
- Marginal Distribution



Probability Review

- Random Variable
- Probability Distribution
- Joint Distribution
- Marginal Distribution

P(Temperature)

T	P
hot	0.5
cold	0.5

P(Weather)

W	P
sun	0.6
rain	0.3
fog	0.1
meteor	0.0

Probability Review

- Random Variable
- Probability Distribution
- Joint Distribution
- Marginal Distribution

P(Temperature)

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Probability Review

- Random Variable
- Probability Distribution
- Joint Distribution
- Marginal Distribution

P(Temperature, Weather)

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

P(Temperature)

T	P
hot	0.5
cold	0.5

P(Weather)

W	P
sun	0.6
rain	0.4

$$p(\text{head}) = 1/2$$

4 queens, 52 total cards

P(queen) = 4/52 = 1/13

Total diamonds = 13

Queen = 1

P(queen/diamond)= 1/13

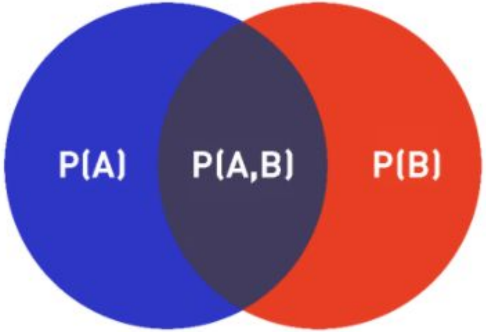
Conditional Probability

P(queen/diamond)= 1/13

P(A/B)= Probability of event A knowing that event B has already occurred

Conditional Probability

$$P(A|B) = P(A, B) / P(B)$$
$$P(A, B) = P(A|B)\, P(B)$$



	select		normalize	
$P(T, W)$	$\rightarrow$	$P(T, W = \text{rain})$	$\rightarrow$	$P(T \mid W = \text{rain})$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

T	W	P
hot	rain	0.1
cold	rain	0.3

T	P
hot	0.25
cold	0.75

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

$$P(queen/diamond) = \frac{P(diamond/queen) * P(queen)}{P(diamond)}$$

P(diamond/queen) = 1/4

P(queen) = 1/13

P(diamond) = 1/4

$$= \frac{1/4 * 1/13}{1/4}$$

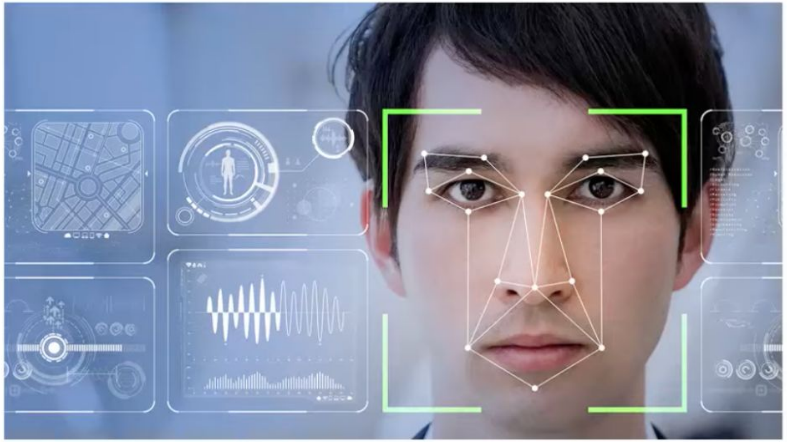
$$= 1/13$$

Bayes Rule: Example

- P(meningitis) = 1/50000
- P(stiff neck) = 1/20
- P(stiff neck | meningitis) = 1/2

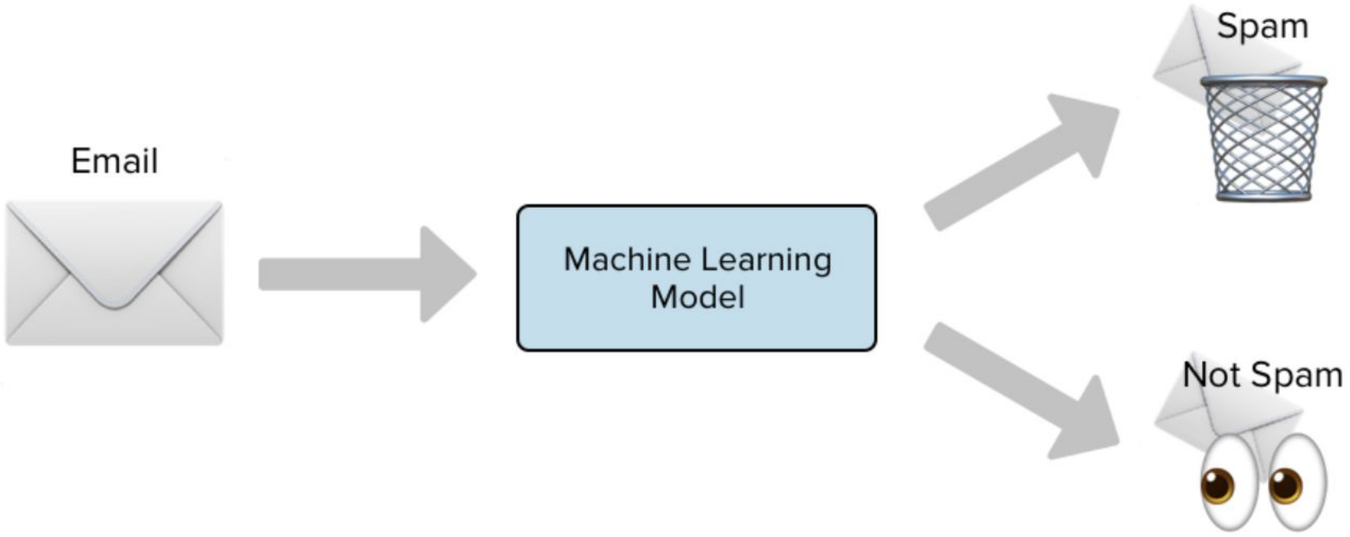
$$P(m|s) = P(m)\,P(S|m) \,/\, P(s)$$
$$= (1/50000)(1/2)/(1/20)$$
$$= (1/5000)$$


SPAM
EMAIL
SPAM




NEWS

Spam Classification



Spam Classification

	class_label	message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...
...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...
5568	ham	Will Ì_ b going to esplanade fr home?
5569	ham	Pity, * was in mood for that. So...any other s...
5570	ham	The guy did some bitching but I acted like i'd...
5571	ham	Rofl. Its true to its name

5572 rows × 2 columns

Spam Classification

Now use a *set* of features:

$$
\begin{aligned}
P(Y|X) &\sim P(Y|F_1, F_2, ..., F_n) \\
&= P(Y, F_1, F_2, ..., F_n) \,/\, P(F_1, F_2, ..., F_n) \\
&= P(Y)\, P(F_1, F_2, ..., F_n|Y) \,/\, P(F_1, F_2, ..., F_n)
\end{aligned}
$$

Here we make the key independence assumption!

$$
\sim \frac{P(Y)\, P(F_1|Y)\, P(F_2|Y) \ldots\, P(F_n|Y)}{P(F_1, F_2, ..., F_n)}
$$

$$
= \frac{P(Y) \prod_i P(F_i|Y)}{\sum_y P(Y) \prod i\, P(F_i|Y)}
$$

Code Review

Smoothing

Laplace Smoothing

- Idea: Pretend we saw every outcome *k* more times than we actually did.
- E.g., if we observed [hht]:
    - MLE: P(heads) = 2/3
    - LAP_1: P(heads) = 3/5
    - LAP_{100}: P(heads) = 102/203
- Smoothing conditionals:
    - LAP_k: $P(x|y) = [count(x,y)+k]/[count(y)+k|X|]$