Naive Bayes *(find and work with a partner)*

Say we have two tests for a specific disease. Each test (features f_1, f_2) can come back either positive "pos" or negative "neg", and the true underlying condition of the patient is represented by y ($y = 1$ is "healthy" and $y = 2$ is "disease"). We observe this training data where $n = 7$ and $p = 2$:

$\boldsymbol{x}$	f_1	f_2	y
$\boldsymbol{x}_1$	pos	neg	1
$\boldsymbol{x}_2$	pos	pos	2
$\boldsymbol{x}_3$	pos	neg	2
$\boldsymbol{x}_4$	neg	neg	1
$\boldsymbol{x}_5$	pos	neg	2
$\boldsymbol{x}_6$	neg	neg	1
$\boldsymbol{x}_7$	neg	pos	2

1. To estimate the probability $p(y = k)$, for $k = 1, 2, \cdots, K$, we will use the formula:

$$\theta_k = \frac{N_k + 1}{n + K}$$

where N_k is the count ("Number") of data points where $y = k$. Compute θ_1 and θ_2. What would θ_1 and θ_2 be if we in fact had *no* training data?

2. To estimate the probabilities $p(x_j = v|y = k)$ for all features j, values v, and class label k, we will use the formula:

$$\theta_{k,j,v} = \frac{N_{k,j,v} + 1}{N_k + |f_j|}$$

where $N_{k,j,v}$ is the count of data points where $y = k$ and $x_j = v$, and $|f_j|$ is the number of possible values that f_j (feature j) can take on. Fill in the following tables with these θ values.

$y = 1$	pos	neg
f_1		
f_2		

$y = 2$	pos	neg
f_1		
f_2		

3. Continuing the example from the previous page, say we have a new data point $\boldsymbol{x}_{\text{test}} = [\text{neg}, \text{pos}]$. Our goal is to predict the class label based on the Naive Bayes posterior probability. In practice, we will compute this probability for each class k, based on our estimates (θ_k and $\theta_{k,j,v}$ terms). Then we will assign this data point the class label with maximum probability:

$$\hat{y} = \underset{k \in \{1,2,\cdots,K\}}{\arg\max}\ p(y=k|\boldsymbol{x}) = \underset{k \in \{1,2,\cdots,K\}}{\arg\max}\ p(y=k)\prod_{j=1}^{p} p(x_j|y=k).$$

For this $\boldsymbol{x}_{\text{test}}$, compute $p(y=1|\boldsymbol{x}_{\text{test}})$ and $p(y=2|\boldsymbol{x}_{\text{test}})$ and then assign a prediction label $\hat{y}$.

4. For the tennis example below, fill in the $\theta_{k,j,v}$ terms (thinking about how this could be implemented using dictionaries).

Day	Outlook	Temperature	Humidity	Wind	PlayTennis (y)
$\boldsymbol{x}_1$	Sunny	Hot	High	Weak	No
$\boldsymbol{x}_2$	Sunny	Hot	High	Strong	No
$\boldsymbol{x}_3$	Overcast	Hot	High	Weak	Yes
$\boldsymbol{x}_4$	Rain	Mild	High	Weak	Yes
$\boldsymbol{x}_5$	Rain	Cool	Normal	Weak	Yes
$\boldsymbol{x}_6$	Rain	Cool	Normal	Strong	No
$\boldsymbol{x}_7$	Overcast	Cool	Normal	Strong	Yes
$\boldsymbol{x}_8$	Sunny	Mild	High	Weak	No
$\boldsymbol{x}_9$	Sunny	Cool	Normal	Weak	Yes
$\boldsymbol{x}_{10}$	Rain	Mild	Normal	Weak	Yes
$\boldsymbol{x}_{11}$	Sunny	Mild	Normal	Strong	Yes
$\boldsymbol{x}_{12}$	Overcast	Mild	High	Strong	Yes
$\boldsymbol{x}_{13}$	Overcast	Hot	Normal	Weak	Yes
$\boldsymbol{x}_{14}$	Rain	Mild	High	Strong	No

y=No (0)

outlook	**Sunny:**	**Overcast:**	**Rain:**
temperature	**Cool:**	**Mild:**	**Hot:**
humidity	**Normal:**	**High:**	
wind	**Weak:**	**Strong:**	

y=Yes (1)

outlook	**Sunny:**	**Overcast:**	**Rain:**
temperature	**Cool:**	**Mild:**	**Hot:**
humidity	**Normal:**	**High:**	
wind	**Weak:**	**Strong:**	