HOMEWORK 3:
WRITTEN EXERCISE PART

1 Multinomial Naïve Bayes [25/2 pts]

Consider the Multinomial Naïve Bayes model. For each point $(\mathbf{x}, y)$, $y \in \{0, 1\}$, $\mathbf{x} = (x_1, x_2, \ldots, x_M)$ where each x_j is an integer from $\{1, 2, \ldots, K\}$ for $1 \leq j \leq M$. Here K and M are two fixed integer.
Suppose we have N data points $\{(\mathbf{x}^{(i)}, y^{(i)}) : 1 \leq i \leq N\}$, generated as follows.

for $i \in \{1, \ldots, N\}$:
  $y^{(i)} \sim$ Bernoulli(ϕ)
  for $j \in \{1, \ldots, M\}$:
    $x_j^{(i)} \sim$ Multinomial$(\theta_{y^{(i)}}, 1)$

Here $\phi \in \mathbb{R}$ and $\theta_k \in \mathbb{R}^K (k \in \{0, 1\}$ are parameters. Note that $\sum_l \theta_{k,l} = 1$ since they are the parameters of a multinomial distribution.
Derive the formula for estimating the parameters ϕ and θ_k, as we have done in the lecture for the Bernoulli Naïve Bayes model. Show the steps.
Solution goes here.

2 Logistic Regression [25/2 pts]

Suppose for each class $i \in \{1, \ldots, K\}$, the class-conditional density $p(\mathbf{x}|y = i)$ is normal with mean $\mu_i \in \mathbb{R}^d$ and identity covariance:

$$p(\mathbf{x}|y = i) = N(\mathbf{x}|\mu_i, \mathbf{I}).$$

Prove that $p(y = i|\mathbf{x})$ is a softmax over a linear transformation of $\mathbf{x}$. Show the steps.
Solution goes here.