New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6923: Written Homework 2.
Due Tuesday, March 7th, 2023, 11:59pm.

Discussion with other students is allowed for this problem set, but solutions must be written-up individually.

Problem 1: Gaussian Naive Bayes (20pts)

In class it was briefly mentioned that the Naive Bayes Classifier can be extended to predictor variables with continuous values (instead of just binary variables). We will derive such an approach here

Consider a data set where each example $(\mathbf{x}, y)$ contains a data vector $\mathbf{x} \in \mathbb{R}^d$ and a label $y \in \{0, 1\}$. As in class, each y is modeled a Bernoulli random variable, which equals 1 with probability p and 0 with probability $1 - p$. To model $\mathbf{x}$ we have two lists of mean/variances pairs:

$$(\mu_{0,1}, \sigma^2_{0,1}), (\mu_{0,2}, \sigma^2_{0,2}), \ldots, (\mu_{0,d}, \sigma^2_{0,d}) \qquad \text{and} \qquad (\mu_{1,1}, \sigma^2_{1,1}), (\mu_{1,2}, \sigma^2_{1,2}), \ldots, (\mu_{1,d}, \sigma^2_{1,d}).$$

If y equals 0, then the j^{th} entry of $\mathbf{x}$ is modeled as an *independent* Gaussian (normal) random variable with mean $\mu_{0,j}$ and variance $\sigma^2_{0,j}$. Alternatively, if y equals 1, then the j^{th} entry of $\mathbf{x}$ is modeled as an independent Gaussian random variable with mean $\mu_{1,j}$ and variance $\sigma^2_{1,j}$.

(a) Given a training data set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ write down mathematical expressions for estimating all model parameters $\mu_{i,j}$ and $\sigma^2_{i,j}$ from the data.

(b) Given a new unlabled predictor vector $\mathbf{x}_{new}$ we would like to predict class label y_{new} using a *maximum a posterior* (MAP) estimate. In other words, we want to choose y_{new} to maximize the posterior probability $p(y_{new} \mid \mathbf{x}_{new})$. Write down an expression for $p(y_{new} \mid \mathbf{x}_{new})$ using Bayes Rule. Write pseudocode for determining if $p(y_{new} = 0 \mid \mathbf{x}_{new})$ or $p(y_{new} = 1 \mid \mathbf{x}_{new})$ is larger. **Hint:** A correct answer should involve the PDF of a Gaussian random variable, and incorporate all model parameters $\mu_{i,j}$ and $\sigma^2_{i,j}$.

(c) If you didn't already in Part (c), modify your pseudocode so that it won't lead to underflow issues when implemented by working with log likelihoods – i.e., your pseudocode should target the problem of determining $\log(p(y_{new} = 0 \mid \mathbf{x}_{new}))$ or $\log(p(y_{new} = 1 \mid \mathbf{x}_{new}))$ is larger.

(d) Implement your method by completing the Python workbook `hw2_stub.ipynb` linked on the course webpage. Attach a printed PDF of your completed notebook results to your homework submission.

Problem 2: Bayesian Central Tendency (12pts)

Let's revisit a question on the first homework from a Bayesian perspective.

(a) Suppose we have a data set of scalar numbers $x_1, \ldots, x_n$. Assume a Bayesian probabilistic model in which the numbers are drawn from a Gaussian distribution with unknown mean μ and variance σ^2. We have no prior information on μ and σ^2: we assume all parameters are equally likely. Prove that the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$ is an MLE estimate for the unknown parameter μ. I.e $\hat{\mu} = \arg\max_{\mu} \Pr(x_1, \ldots, x_n \mid \mu)$.

(b) Now assume a Bayesian probabilistic model in which the numbers are drawn from a Laplace Distribution with unknown mean μ and variance $2b^2$. Prove that the sample median is a MLE estimate for the unknown parameter μ.

(c) Suppose $\mu \in [0, 1]$ and $x_1, \ldots, x_n$ are drawn i.i.d from a Bernoulli distribution with parameter μ. I.e. x_i is 1 with probability μ and 0 with probability $1 - \mu$. Prove that the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$ is also an MLE estimator for μ in this setting.