CSE 151: Machine learning **Winter 2019**

Homework 1

Submission instructions:

- If a problem asks for a numerical answer, you need only provide this answer. There is no need to show your work, unless you would like to.
- Please type up your solutions. We suggest using an online latex editor like `www.overleaf.com`, though this is not a requirement.
- Upload the PDF file for your homework to `gradescope` by midnight on Thursday Feb 7.

1. For each of the following prediction tasks, state whether it is best thought of as a *classification* problem or a *regression* problem.
   (a) Based on sensors in a person's cell phone, predict whether they are walking, sitting, or running.
   (b) Based on sensors in a moving car, predict the speed of the car directly in front.
   (c) Based on a student's high-school SAT score, predict their GPA during freshman year of college.
   (d) Based on a student's high-school SAT score, predict whether or not they will complete college.
2. Two fair dice are rolled. What is the probability that:
   (a) Their sum is 10, given that the first roll is a 6?
   (b) Their sum is 10, given that the first roll is an even number?
   (c) They have the same value?
3. A certain genetic disease occurs in 5% of men but just 1% of women. Let's say there are an equal number of men and women in the world. A person is picked at random and found to possess the disease. What is the probability, given this information, that the person is male?
4. A fair die is rolled twice. Let X_1 and X_2 denote the outcomes, and define random variable X to be the minimum of X_1 and X_2.
   (a) Determine the distribution of X.
   (b) What is $\mathbb{E}(X)$?
   (c) What are the variance and standard deviation of X?
5. In each of the following cases, say whether X and Y are independent.
   (a) Pick a random sentence out of *Hamlet*. X is the first word in the sentence, Y is the second word.
   (b) Randomly pick a card from a pack of 52 cards. Define X to be 1 if the card is a Jack, and 0 otherwise. Define Y to be 1 if the card is a spade, and 0 otherwise.
   (c) Randomly pick two cards from a pack of 52 cards. X is 1 if the first card is a spade, and 0 otherwise. Y is 1 if the second card is a spade, and 0 otherwise.

6. Would you expect the following pairs of random variables to be uncorrelated, positively correlated, or negatively correlated?
   (a) The amount of rainfall on a given day and the amount of rainfall the following day.
   (b) The number of people at the beach on a given day and the number of people skiing that day.
   (c) A person's age and social security number.

7. Each of the following scenarios describes a joint distribution (x, y). In each case, give the parameters of the (unique) bivariate Gaussian that satisfies these properties.
   (a) x has mean 2 and standard deviation 1, y has mean 2 and standard deviation 0.5, and the correlation between x and y is -0.5.
   (b) x has mean 1 and standard deviation 1, and y is equal to x.

8. Roughly sketch the shapes of the following Gaussians $N(\mu, \Sigma)$. You only need to show a representative contour line which is qualitatively accurate (has approximately the right orientation, for instance).
   (a) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$
   (b) $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & -0.75 \\ -0.75 & 1 \end{pmatrix}$

9. For each of the two Gaussians in the previous problem, check your answer using Python: draw 100 random samples from that Gaussian and plot them.

10. In this problem, you are going to build a generative digit classifier using the same MNIST dataset you used for nearest neighbors classification in Homework 0. If you don't already have them, first go and download the train, validation, and test sets from the class website. (For a description of the data format, please see the last problem of Homework 0.)

    The models that you will be developing will be parameterized in terms of several component distributions: (1) a multinomial distribution that specifies the prior probability of each digit label, $Y \in \{0, 1, 2, \ldots 9\}$, and (2) a set of Gaussian distributions, one for each digit label, that specify the conditional probability of the input observation, $X \in \mathbb{R}^d$, given the digit label, Y. The generative process your model assumes is as follows: First, sample a label Y from the prior. Next, sample an observation X from the Gaussian corresponding to your sampled label Y.

    We will denote the parameters of the multinomial prior as the vector π, where $P_\pi(Y = y) = \pi_y$. We will denote the parameters of the Gaussian corresponding to digit label y as (μ_y, Σ_y). In the first parts of this problem, we will be dealing with univariate Gaussians with scalar parameters. In later parts, we will consider multivariate Gaussians with vector-valued μ_y and matrix-valued Σ_y.

    (a) First, you are going to implement a simple generative classifier that you will learn from the training data using maximum likelihood estimation (MLE). For this part, let your observation X be *the sum of the values of all pixels* in the raw input image. This means that instead of observing a high-dimensional vector specifying all the pixels in the digit image, your model will only observe a single scalar representing the total amount of light in the image. This means that $X \in \mathbb{R}$ and that your Gaussians will be univariate. What are the MLE parameter estimates $\hat{\pi}$, $\hat{\mu}_0, \ldots, \hat{\mu}_9$, and $\hat{\Sigma}_0, \ldots, \hat{\Sigma}_9$ on the training set? [Hint: you will need to write code to compute empirical counts, means, and variances from the training data.]

(b) What is the validation error of the univariate classifier you just trained? [Hint: You will need to apply Bayes' rule to compute the posterior probability $P(Y|X)$ under the model you just learned. Also, don't worry if your error rate is quite high – this is expected.]

(c) The error rate of your simple classifier should be relatively high. A single feature is not sufficient to discriminate between all nine classes. As we saw in the winery problem in lecture, adding additional features to your observation space can improve performance.

For this problem, we want you to try to improve the performance of your classifier by incorporating at least one new feature into your observation space. You can choose whatever additional feature(s) you think might be useful. Note, your model will have to change to accommodate multi-dimensional observations: the Gaussian emission distributions will become multivariate, with dimensionality matching your observation space. This means your observations (and Gaussians) must be at least two-dimensional.

Feel free to explore as many new features as you like. In order to choose the best combination of features, feel free to tune using validation error. In the end, report the test error of your best feature combination and briefly describe the new feature(s) you introduced.