ECE 492-45 Homework 5 (Fall 2021)
Instructor: Dr. Chau-Wai Wong
Material Covered: LSTM, Gradient Descent, Statistical Learning Basics

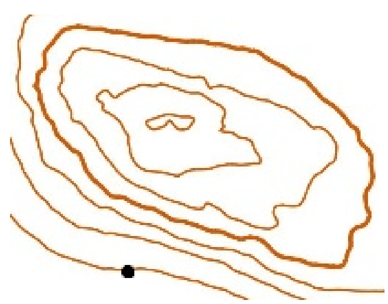
Problem 1 (20 points) [Character-level LSTM] In *this Colab notebook file*, you will use a recurrent neural network with long short-term memory (LSTM) units to predict the next character based on a Shakespeare writing. The trained model auto-generates texts, which can imitate the writing style of Shakespeare. To start text generation, you should pass a starting char(acter), from which you then generate one character at a time. Examine the following items:

1. Test the trained model by passing different starting strings. Also, train with your own dataset, e.g., writing from another author, and show the results.
2. Use one sentence each to explain what the following functions do: `random_chunk()` and `random_training_set()`.
3. Draw an unrolled block diagram for the following code segment:

```
for c in range(chunk_len - 1):
    out_target = target[c].unsqueeze(0).type(torch.LongTensor)
    out, hidden, cell = model(inp[c], hidden, cell)
    loss += criterion(out, out_target)
```

Problem 2 (20 points) [Level Curves and Gradient Descent]

a) A set of level curves is shown as follows. Use the dot as the starting point, draw a trajectory of gradient descent steps. Annotate each descent step using a line segment with an arrow at the end. Explicitly draw a tangent line at each step, which can assist you to determine the negative gradient direction. You may vary the descent step size.



b) A cost function is given as $J(w_1, w_2) = \exp(-w_1^2 - 10w_2^2)$. Use the command you learned from Problem 2 of HW1 to draw a contour plot/set of level curves. Illustrate by drawing two initial points and their gradient descent trajectories using a fixed step size. One initial point must lead to fast convergence and the other must lead to slow convergence. (You may draw the trajectories on a printout, or you may do a screenshot, paste it into Powerpoint, draw the arrows and tangent lines, and save it as a PDF file.)

Problem 3 (20 points) [Conditional Expectation, Variance Operator]

a) Given the joint PMF for random variables X and Y in the table, compute the following quanti-

Table 1: Joint PMF, $p_{XY}(x, y)$

$Y \setminus X$	1	2	3
0	0.3	0.1	0.3
1	0.1	0.1	0.1

ties and tabulate your results: $p_X(x)$, $p_Y(y)$, $p_{X|Y}(x|y)$, $p_{Y|X}(y|x)$, $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathbb{E}[X|Y = y]$, $\mathbb{E}[Y|X = x]$. (Intermediate steps must be shown to receive full points.) Explain the difference between $\mathbb{E}[X|Y = y]$ and $\mathbb{E}[X|Y]$.

b) Prove the following formulas:

$$\text{Var}(aX + b) = a^2 \text{ Var}(X), \tag{1a}$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{ Cov}(X, Y), \tag{1b}$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y), \text{ when } X \text{ and } Y \text{ are uncorrelated.} \tag{1c}$$

$$\text{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \text{ Var}(X_i),\ X_i\text{'s uncorrelated. Useful for Problem 3e.} \tag{1d}$$

You may find the following equations useful: i) the shortcut formulas for variance, $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$; and ii) the covariance, $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. Answer the following questions:

- Why does b not appear on the right-hand side of (1a)?
- How does the variance of sum of two random variable compare to the sum of the variance of individual variables when the variables are negatively/anti-correlated? Can you give an extreme example?
- Why is it a plus sign rather than a minus sign on the right-hand side of (1c)?

Problem 4 (20 points) (20 points, try it after Wed. lecture) [Curse of Dimensionality] Read the first paragraph of the problem statement of *ESLII-2.4*. Note that we may also write $\mathbf{X} = (X_1, X_2, \ldots, X_p)$, where $X_k \sim \mathcal{N}(0, 1)$ for $k = 1, \ldots, p$. Use a programming language of your choice. To get started, set $p = 10$. Note that in this problem, all vectors are column vectors.

a) Write a computer program to randomly draw/generate $N = 100$ vectors from the template random vector $\mathbf{X}$, namely, $\{\mathbf{x}^{(i)},\ i = 1, \ldots, N\}$. Note that each vector should contain p normality distributed random numbers. Plot all vectors as points in a 3-D space consisting of the first, second, the last coordinates.

b) Calculate the coordinate value of each point after being projected on to a fixed direction specified by $\mathbf{a} = \mathbf{x}_0/||\mathbf{x}_0||$, namely, $z^{(i)} = \mathbf{a}^T\mathbf{x}^{(i)}$. Here, $\mathbf{x}_0$ is an arbitrary nonzero vector of length p, "T" is the transpose operation, and $z^{(i)} \in \mathbb{R}$. What are the sample mean and sample variance of the projected coordinates $\{z^{(i)},\ i = 1, \ldots, N\}$?

c) Repeat a) and b) for $p \in [1, 80]$. You may want to use a `for` loop to achieve this. Optionally, put your code for parts a) and b) into a function to make your code easier to read. Plot the sample variance of the projected coordinates as a function of p.

d) Calculate the squared distance of each point to the origin, namely, $d_i^2 = ||\mathbf{x}^{(i)}||^2$. What is the sample mean of $\{d_i^2,\ i = 1, \ldots, N\}$? Plot the sample mean of the squared distance as a function of p in the same plot of c). Limit the range of y-axis between 0 and 80. For $p = 5$, inspect the values of any five d_i^2's. Do the results in b) and c) match with conclusion drawn in the third paragraph of *ESLII-2.4*?

e) Use the formulas from Problem 1b, prove that $\text{Var}(Z) = 1$ where $Z = \mathbf{a}^T\mathbf{X}$, and $\mathbb{E}[D^2] = p$ where $D = ||\mathbf{X}||$. Are the theoretical results in this part consistent with the simulated results obtained in c) and d)?

Problem 5 (20 points) [`Auto` Data Analysis] Complete *ISLR-2.4.9*.

For Python users, please follow the text book's instructions while referring to *the "equivalence" Python code*, where you may find the sample code and the comments useful.