ECE 11 Homework 5 (Fall 2022)
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 char at a time. Examine the following items:

a) Test the trained model by passing different starting chars. Also, train with your own dataset, e.g., writing from another author, and show the results.

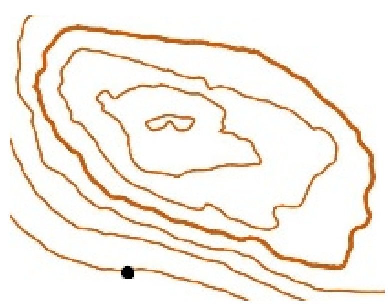
b) Use one sentence each to explain what the following functions do: `random_chunk()` and `random_training_set()`.

c) 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 HW2 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.)

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

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

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 quantities 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:

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

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

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

$$\mathrm{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2\ \mathrm{Var}(X_i),\ X_i\text{'s uncorrelated. Useful for Problem 4e.} \tag{1d}$$

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

a) We are given two variables X and Y that are not independent. Hence, we may use one to estimate the other. Find the best deterministic function $g(\cdot)$ such that it minimizes the expected squared error between Y and $g(X)$ conditioned on $X=x$. You may find a change of variable using θ in the place of $g(x)$ helpful. Pay attention to write clearly the upper case X and the lower case x in your submission.

b) *Arithmetic average*, or the *sample mean* in a statistical context, is commonly used in everyday life for making quantitative description. We examine a statistical interpretation for the arithmetic average below. A person weighs μ lb. He tried multiple scales in a supermarket and recorded the reading from each scale, denoted by Y_i for the ith scale. We may create a linear model as follows to relate the true weight μ and the measurement Y_i:

$$Y_i = \mu + e_i, \quad i = 1, \ldots, N,$$

where e_i is the measurement error of the ith scale. Use the mean-square criterion $J(\mu) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \mu)^2$ to find the closed-form expression for the best estimator for μ. The expression should contain $\{Y_i\}_{i=1}^{N}$ only, and should not contain such symbols as μ or e_i as they were not available when readings were recorded. Does the expression make intuitive sense?

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.