Quiz 1	**Name:** _____
Week 1, Sep/7/2022	**On your left:** _____
CS 280: Fall 2022	**On your right:** _____
Instructor: Lan Xu	

Instructions:
Please answer the questions below. Show all your work. . NO discussion/collaboration is allowed.

Problem 1. (10 points) **variance and Conditional probability**.

(a) Assume $x_1, \cdots, x_n$ are independent random variables, show that $var(x_1 + x_2... + x_n) = var(x_1) + var(x_2) + ...var(x_n)$

(b) C= $\{C_1, C_2...C_n\}$ show that $P(C_i|A, B) = \frac{P(C_i,B|A)}{\sum_i P(C_i,B|A)}$

Problem 2. (10 points) *Learning basics.* Consider Ridge regression

$$L(\mathbf{w}) = \sum_{i=1}^{n} ||\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - y_i||^2 + \lambda \mathbf{w}^{\mathrm{T}}\mathbf{w}$$

(a) Show that its optimal weight $\mathbf{w}^* = (X^{\mathrm{T}}X + \lambda I)^{-1}X^{\mathrm{T}}Y$ with properly defined X and Y.

(b) (bonus 10 points) Provide a probabilistic formulation of the objective function using the Bayesian Theorem.

Problem 3. (10 points) *Gradient.* Let $\sigma(a) = \frac{1}{1+e^{-a}}$ be an activation function, and $f(\mathbf{w}) = -\sum_{i=1}^{n}[y_i \log(\mu_i) + (1-y_i)\log(1-\mu_i)]$, where $\mu_i = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i)$.

(a) Compute $\frac{d}{d\mathbf{w}} f(\mathbf{w})$.

(b) (bonus 10 points) Show that $f(\mathbf{w})$ is convex.

Problem 4. (10 points) *Python Programming*. Consider estimating the model parameter $\mathbf{w}$ in **Problem 3** through minimizing $f(\mathbf{w})$. We will use gradient descent to achieve this. Specifically, assume we have a set of data $X = [\mathbf{x}_i]_{i=1}^n$ and their corresponding target $Y = [\mathbf{y}_i]_{i=1}^n$. Our predictor is defined by $\sigma(a)$ and the loss function is defined by $f(\mathbf{w})$ as in **Problem 3**.

Below we have provided an outline and you need to fill in the missing part.

```
import numpy as np

N, D_in, D_out = 64, 1000, 1

# random data generation
x = np.random.randn(N, D_in)
y = np.random.randn(N, D_out)

# randomly initialize weights
w = (____________________________)

# set up learning rate
lr = 1e-6

# iterative gradient descent
for it in range(500):
    # forward pass
    h = (____________________________)
    y_pred = (____________________________)

    # compute loss
    loss = (____________________________)

    # compute gradient
    grad_y_pred = (____________________________)
    grad_w = (____________________________)

    # update w
    (____________________________)

# (Bonus 5 points): Write test procedure below
```