Course: Machine Learning Techniques
Week 2: Practice Questions

Practice questions

1. (1 point) Small learning rate makes convergence in gradient descent algorithm
   - A. slower
   - B. faster
   - C. has no effect
   - D. impossible

   Answer: A

2. (1 point) If we have **p** training examples each having **f** features, then the shape of the feature matrix is
   - A. $p \times f$
   - B. $f \times p$
   - C. $(p-1) \times (f-1)$
   - D. $(f-1) \times (p-1)$

   Answer: A

3. (1 point) Following table shows the values of actual and predicted labels. Find the sum squared error.

Actual	Predicted
1	1.8
2	2.4
3	3.6
4	4.2

   - A. 0.55
   - B. 0.50
   - C. 0.45
   - D. 0.40

   Answer: A

4. (1 point) The gradient of $f(x_1, x_2) = x_1^2 + x_2^2$ at (0,0) is
   - A. $[1, 1]^T$

B. $[0, 0]^T$

C. $[1, -1]^T$

D. $[-1, -1]^T$

Answer: B

consider the following code

```
def predict(X,w):
    assert X.shape[-1]==w.shape[0], "X and w don't have compatible    dimer
    return X @ w

def gen_data(n):
    #obtain y=6+5x+noise
    w0=6
    w1=5
    X=10*np.random.rand(2*n,)
    y=w0+w1*X+np.random.randn(2*n,)
    return X,y

X,y=gen_data(200)
```

5. (points) To generate 500 training data points randomly the value of parameter in gen_data(n) should be?

   A. 500

   B. 250

   C. 200

   D. 100

   Answer: B

6. (points) For fairly large number of training examples the time of execution for vectorized form as compared to non vectorized form is

   A. more

   B. less

   C. same

   Answer: B

7. (1 point) (Multiple select)Which of the following are hyper parameters in gradient descent algorithm?

   A. number of features

   B. no. of iterations

C. learning rate

D. loss function

Answer: B,C

8. (1 point) Consider the following model:
$y = 3x_1 + 4x_2 + 5$
The number of parameters and features in the above model are respectively

A. 2,3

B. 3,2

C. 3,3

D. 1,1

Answer: B