Homework 3: Machine Learning Basics

Due Mar 14 at 11:59pm **Points** 100 **Questions** 21
Available Jan 30 at 12am - May 1 at 11:59pm 3 months **Time Limit** None
Allowed Attempts 200

Instructions

Submission later than the due will be penalized. 2% will be deducted per 24 hours after the due.

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	10,449 minutes	81 out of 100

(!) Answers will be shown after your last attempt

Score for this attempt: **81** out of 100
Submitted Feb 21 at 11:39pm
This attempt took 10,449 minutes.

Question 1

5 / 5 pts

For a polynomial regression method, if you increase the degree of the polynomial from p=3 to p=5, then the training error and test error both decreases. It indicates the model with p=3 suffers from

A. Underfitting

B. Overfitting

C. Neither A or B

● A

○ B

○ C

Question 2

5 / 5 pts

For a polynomial regression method, if you increase the degree of the polynomial, the training error decreases but the test error increases. As you increase the degree, what is happening to the model?

A. Underfitting

B. Overfitting

C. Neither A or B

○ A

● B

○ C

Question 3

5 / 5 pts

Consider the ridge regression model: $\min_{\mathbf{w}} \ \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2$. Let $\mathbf{H}$ be the Hessian matrix (aka the second derivative) of the objective function.

If you decrease λ, then how will the condition number of the $\mathbf{H}$ change?

(Hint: condition number is the max eigenvalue over the min eigenvalue.)

● It will increase.

○ It will decrease.

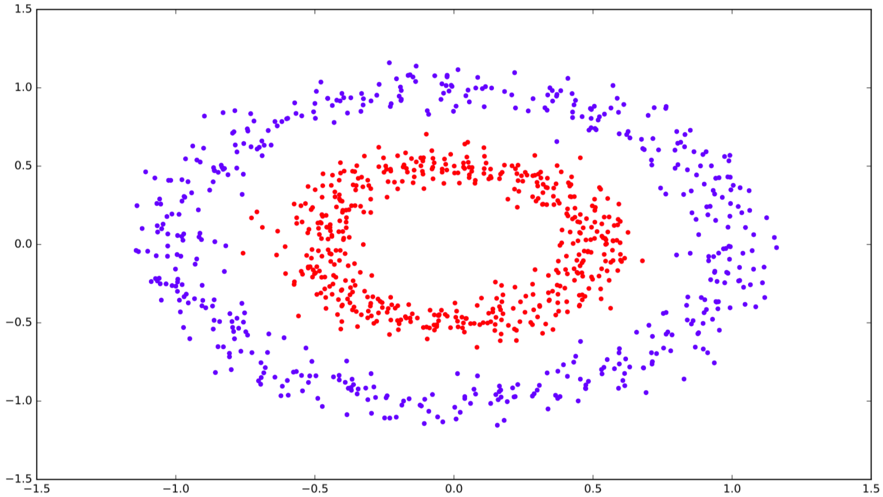
○ It will not change.

Question 4

4 / 4 pts

Let the 2-dim feature vectors be scattered in the following plot. The colors indicate the class labels.

Which method has the poorest classification accuracy on this dataset?



○ A. Kernel support vector machine.

● B. Logistic regression.

○ C. K-nearest-neighbor (KNN) classifier.

○ D. A 2-layer full-connected neural network.

Question 5

4 / 4 pts

Which of the following is NOT a classification method?

○ A. Kernel support vector machine.

○ B. Multi-layer perceptron.

○ C. Softmax classifier.

○ D. Decision tree.

● E. Principal component analysis.

○ F. Naive Bayes.

Question 6

4 / 4 pts

Dimensionality reduction is a family of unsupervised learning tasks.

○ True

● False

Question 7

5 / 5 pts

The set $C = \left\{ \mathbf{x} \in \mathbb{R}^3 \ : \ \|\mathbf{x}\|_{10} \leq 100 \right\}$ is a convex set. (Here, $\|\mathbf{x}\|_p$ denotes the ℓ_p-vector norm.)

● True

○ False

Question 8

5 / 5 pts

You are required to build a convolutional neural network to solve a hand-written character recognition problem. Which of the following is the best choice for the output layer?

A. Softmax function.

B. Logistic function.

C. ReLU.

D. Max Pooling.

● A

○ B

○ C

○ D

Incorrect

Question 9

0 / 5 pts

You are required to predict people's age based on 256×256 photos. You train a deep neural network using $n = 1,000,000$ samples.

You use your model to make prediction for a batch of 64 test samples; what is the shape of the output of the neural network?

A. 1×64

B. 2×64

C. 64×64

D. 256×64

E. $1,000,000 \times 64$

F. 256×256

G. $256 \times 256 \times 64$

○ A

○ B

○ C

○ C

○ E

○ F

● G

The is a regression problem. So the prediction for an input photo is a scalar.

Question 10

5 / 5 pts

You train a deep convolutional neural network for handwritten digit recognition using the *accelerated gradient descent algorithm*. The objective function value and the training error do not change after the1,000th step.

After 100,000 steps, the algorithm will reach a local minimum.

○ True

● False

Gradient descent and accelerated gradient descent easily get stuck in a saddle point. They are extremely unlikely to reach a local minimum.

Incorrect

Question 11

0 / 5 pts

You train a deep convolutional neural network using stochastic gradient descent. At the 200th iteration, you evaluated the full gradient and found it exactly zero.

At the 200th iteration, the algorithm reached a local minimum.

● True

It can be a local minimum or a saddle point.

You must also check the smallest eigenvalue of the Hessian matrix. If it is a local minimum, then the Hessian matrix at this point is positive semidefinite.

○ False

Question 12

5 / 5 pts

You want to train a deep convolutional neural network for object recognition. Which of the following does NOT typically improve the prediction accuracy on the test set?

A. Pretrain the model on a large dataset.

B. Use dropout as a regularization.

C. Increase the number of GPUs.

D. Use data augmentation.

○ A

○ B

● C

○ D

Incorrect

Question 13

0 / 5 pts

You want to train a deep convolutional neural network on *a small dataset* for object recognition. The number of training samples is far smaller than the number of network parameters.

What of the following can **improve the prediction accuracy** on the test set?

A. Pretrain the neural network on ImageNet; then fix the bottom layers and train the top layers (including the output layer) on the small dataset.

B. Pretrain the neural network on ImageNet; then fix the top layers and train the bottom layers (including the input layer) on the small dataset.

C. If the algorithm is mini-batch stochastic gradient descent, then use a larger batch size.

D. If the algorithm is mini-batch stochastic gradient descent, then use a smaller batch size.

○ A

● B

Please go to the 2nd lecture on CNN.

○ C

○ D

Question 14

4 / 4 pts

Match the activation functions and their names.

A. $\sigma\left(z\right)=\frac{1}{1+e^{-z}}$

B. $\sigma(z)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$

C. $\sigma(z)=\begin{cases} z, & \text{if } z\geq 0; \\ 0, & \text{otherwise.} \end{cases}$

D. $\sigma(z)=\begin{cases} z, & \text{if } z\geq 0; \\ 0.01z, & \text{otherwise.} \end{cases}$

Sigmoid	A
Tanh	B
ReLU	C
Leaky ReLU	

D

Question 15

5 / 5 pts

Let $\sigma\left(z\right) = \frac{1}{1+e^{-z}}$.

The lower bound on $\sigma(z)$ is $\inf_z \sigma(z)$ = 0

The upper bound on $\sigma(z)$ is $\sup_z \sigma(z)$ = 1

(Hint: the answers are integers.)

Answer 1:

0

Answer 2:

1

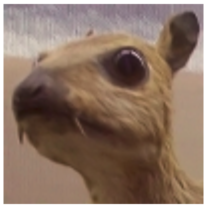
Question 16

5 / 5 pts

We apply the filter (kernel)

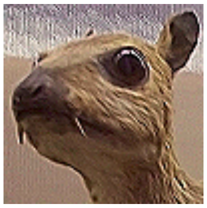
$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

to the following image:

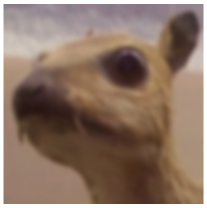


Which of the following is most likely the result:

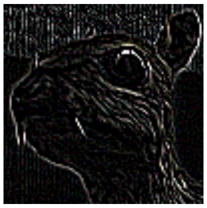
A.



B.



C.



○ A

○ B

● C

The filter detects edges.

Question 17

5 / 5 pts

Let $\{v_1,\ v_2,\ v_3\}$ be an orthonormal basis, $\sigma_1 = 10,\ \sigma_2 = 5,\ \sigma_3 = 1$, and $A = \sum_{i=1}^{3} \sigma_i v_i v_i^T$ be a matrix.

What is the squared $\ell_2 -$norm: $||A\ v_2||_2^2$?

Hint:

- Orthonormal basis has 2 properties: (1) unit L2 norm and (2) orthogonal to each other.
- The matrix-vector product, $A\ v_2$, is a vector.

○ A. 1

○ B. 5

○ C. 10

○ D. 16

● E. 25

○ F. 100

○ E. 126

Incorrect

Question 18

0 / 4 pts

The Level 3 BLAS (BLAS3) perform matrix decompositions, e.g., QR decomposition and SVD.

● True

LAPACK performs matrix inversion and matrix decompositions. LAPACK and BLAS3 do not have overlap.

○ False

Question 19

5 / 5 pts

Replacing for-loops by equivalent Numpy built-in matrix and vector functions always improves efficiency.

● True

○ False

Question 20

5 / 5 pts

You want to train a logistic regression model using a dataset with 10,000 samples. The dataset has 500 positive samples (i.e., y=+1) and 9,500 negative samples (i.e., y=-1).

Evaluating the training, validation, or test performance by the **ROC curve** is a good idea.

● True

○ False

○ True

○ False

Question 21

5 / 5 pts

The quality of wine can be (1) outstanding, (2) very good, (3) good, (4) mediocre, or (5) not recommended. We want to predict the quality based on the measurements of chemicals, e.g., alcohol, malic acid, magnesium, etc.

(Hint: outstanding'' is better than very good'', very good'' is better than good', and so on.)

○ This is a classification task.

● This is a regression task.

○ This is a clustering task.

○ This is a dimensionality reduction task.

Quiz Score: **81** out of 100