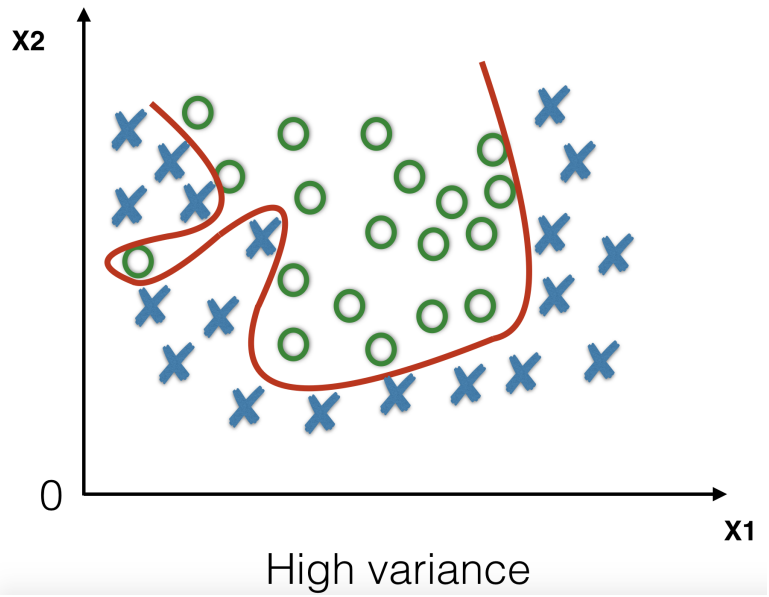
deeplearning.ai

Machine Learning Engineers Test - (1h45)

23m
to test end

## ☆ Variance

You are building a model to separate the data below. Green circles (resp. blue crosses) indicate positive (resp. negative) labels.




Fill-in the blank: Based on the graph above, we can say that the model has _____ variance.

Pick one of the choices

- low
- appropriate
- high

Clear selection

## ☆ PCA: Number of principal components

You've run the Principal Components Analysis algorithm on your data. You obtained the following eigenvalues: 2.421, 1.880, 1.672, 0.962, 0.0031, 0.0029, 0.00093.

How many principal components would you keep?

Pick one of the choices

- (•) 3
- ( ) 4
- ( ) 5
- ( ) 6
- ( ) 7

Clear selection

## ☆ Support Vector Machine

In Support Vector Machines, kernel functions map low dimensional data into a high dimensional space. True/False.

Pick one of the choices

- (•) True
- ( ) False

Clear selection

## ☆ Loss functions

You are given the following dataset:

| height | speed | time spent | category (0/1) |
|---|---|---|---|
| 183 | 8.3 | 14.2 | 1 |
| 165 | 9.4 | 12.9 | 0 |
| 192 | 9,2 | 15.5 | 1 |
| ... | ... | ... | ... |

The features are height, speed and time spent; the label you want to predict is a binary category.

What loss would you choose to minimize in order to train a model to solve this task?

A) $L = -y \log(\hat{y}) - (1-y) \log(1-\hat{y})$
B) $L = y \log(\hat{y}) + (1-y) \log(1-\hat{y})$
C) $L = || \ y - \hat{y} \ ||_2^2$
D) $L = || \ y - \hat{y} \ ||_2^2 + constant$

Pick one of the choices

- (•) A
- ( ) B
- ( ) C
- ( ) D

Clear selection

## ☆ L2 Regularization

Using L2 regularization will introduce sparsity in your weight parameters.

Pick one of the choices

- ( ) TRUE

FALSE

Clear selection

## ☆ Bayes error

In the validation set, the Bayes error is lower than the human-level error.

Pick one of the choices

TRUE

FALSE

Clear selection

## ☆ Gradient Descent Optimization

Let L be the following cost function:

$$L = f(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

where $h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$ and $x^{(i)}, y^{(i)}$ are scalars.

In this setting, what is the update rule of a Gradient Descent Optimizer?

A. $\begin{cases} \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \\ \theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} \end{cases}$

B. $\begin{cases} \theta_0 := \theta_0 (1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}) \\ \theta_1 := \theta_1 (1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}) \end{cases}$

C. $\begin{cases} \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} \\ \theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} \end{cases}$

D. $\begin{cases} \theta_0 := \theta_0 (1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})) \\ \theta_1 := \theta_1 (1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}) \end{cases}$

Pick one of the choices

- (•) A
- ( ) B
- ( ) C
- ( ) D

Clear selection

## ☆ Carrying out error analysis

You've trained a model on 1,000,000 images for a multi-class classification task. You are carrying out error analysis and counting up what errors the algorithm makes on the dev set.

Which of these datasets do you think you should manually go through and carefully examine, one image at a time?

Pick one of the choices

- ( ) 10,000 randomly chosen images
- (•) 10,000 images on which the algorithm made a mistake
- ( ) 500 randomly chosen images
- ( ) 500 images on which the algorithm made a mistake

Clear selection

## ☆ Human-level performance

After working for a year, you finally achieve:

| | |
|---|---|
| Human-level performance | 0.20% |
| Training set error | 0.10% |
| Dev set error | 0.10% |

What can you conclude? (Check all that apply.)

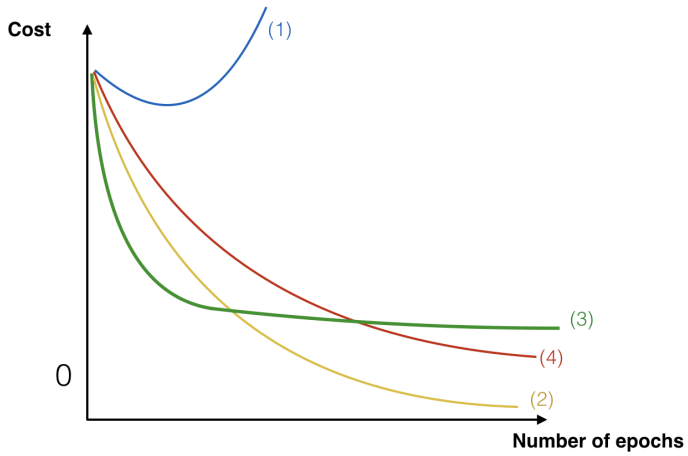
Pick the correct choices

- [ ] **This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.**
- [x] **If the test set is big enough for the 0.10% error to be accurate, this implies Bayes error is less than or equal to 0.10%**
- [x] **It is now harder to measure avoidable bias, thus progress will be slower going forward.**

Clear selection

## ☆ Hyperparameter tuning: Learning rate.

During the hyperparameter tuning phase, you train the same model with 4 different learning rates. You plot the loss functions for each of these learning rates and obtain the graph below. How would you characterize the learning rate used in experiment (3)? (Assume each of the 4 possible answers correspond to one experiment's learning rate.)




Pick one of the choices

- low
- very high
- high
- good

Clear selection

## ☆ Data split

You are working on a binary classification project to classify the presence of cats in images. They give you 1,000,000 images.

Before implementing your algorithm, you need to split your data into train/dev/test sets. Which of these do you think is the best choice?

Pick one of the choices

- Train set: 600,000 Dev set: 300,000 Test set: 100,000
- Train set: 950,000 Dev set: 25,000 Test set: 25,000
- Train set: 600,000 Dev set: 100,000 Test set: 300,000
- Train set: 333,334 Dev set: 333,333 Test set: 333,333

Clear selection

## ☆ Improving performance on the training set

You train a system, and its errors are as follows (error = 100% - accuracy):

| | |
|---|---|
| Training set error | 5.0% |
| Dev set error | 5.5% |

This suggests that one good avenue for improving performance is to train a bigger network so as to drive down the 5.0% training error. Do you agree?

Pick one of the choices

- **Yes, because having 5.0% training error shows you have high avoidable bias.**
- **Yes, because this shows your avoidable bias is higher than your variance.**
- **No, because this shows your variance is higher than your avoidable bias.**
- **No, because there is insufficient information to tell.**

Clear selection

## ☆ L1 and L2 loss functions

You are trying to achieve a linear regression task using machine learning. Which of the following statements about the L1 and L2 loss functions is true? (check all that apply.)

Reminder:

- The L1 loss function is defined by $|y-\hat{y}|$ where y is the ground truth label and yhat is the output prediction.
- The L2 loss function is defined by $|y-\hat{y}|^2$ where y is the ground truth label and yhat is the output prediction.

Pick the correct choices

- [ ] **L2 is more robust to outliers.**
- [ ] **L1 has possibly multiple solutions.**
- [x] **The L2 loss function has analytical solution**

Clear selection

## ☆ K-means

Which of these hyperparameters (or choices) are inputs to the K-means algorithm? (Check all that apply.)

Pick the correct choices

- [x] Number of clusters
- [ ] The activation function to be used
- [x] The centroids' update rule
- [ ] Number of components retained

Clear selection

## ☆ ML: Model accuracy

Our task is to build a trigger word detector. This system should be able to detect a target word from audio clips. One input example is 1000 time-step audio recording, and a binary label is assigned to each time-step:

- 0 indicates that the trigger word hasn't been said in the past 10 time steps.
- 1 indicates that the trigger word has been said in the past 10 time steps.

We define accuracy by the ratio of correctly predicted time steps over the total number of time steps. On average there the trigger word is said 3 times in 1000 time-steps. Your model performs 97% accuracy when tested on a large number of audio recordings.

What would you say about the model?

Pick one of the choices

- ( ) It is a very good model, and it will be hard to do a lot better.
- (•) It seems that the model works but it could probably do better.
- ( ) It is a bad model because the accuracy is not high enough

- There is not enough information to tell

Clear selection

## ☆ K-Nearest Neighbors

Which of the following properties of the K-Nearest Neighbors (kNN) classifier is **false**?

Pick one of the choices

- kNN doesn't make any assumptions about data
- The higher is the dimension of the data, the more stable is the kNN algorithm.
- It is recommended to normalize the dataset before using kNN, because features with different scales and variances can significantly hurt the performance of kNN.
- kNN can be used for multi-class classification problems.

Clear selection

## ☆ k-Nearest Neighbors

Which, if any, of the following properties of the k-Nearest Neighbors (kNN) classifier is **true**? (check all that apply.)

Pick the correct choices

- [x] kNN falls in the family of unsupervised learning algorithms.
- [x] kNN is generally slower predicting the label of a given example compared to a shallow neural network (ANN).
- [ ] kNN is a parametric learning algorithm.
- [ ] kNN can't use similarity metrics such as the Manhattan or the Euclidean distance.

Clear selection

## ☆ Data split in Healthcare

You are working with a group of doctors to solve one of their image classification problem with machine learning. They have collected 100,000 images from microscopes and gave them to you. Images have been taken from three types of microscopes:

- A (50,000 images)
- B (25,000 images)
- C (25,000 images).

The doctors who hired you would like to use your algorithm on images from microscope C. Which of the following propositions are true? (Check all that apply.)

Pick the correct choices

- [x] **There should be only images from C in the test set.**
- [x] **There should be only images from C in the dev set.**
- [x] **There should be only images of C in the train set.**
- [x] **It would be reasonable to have a train/dev/test data split of roughly 60%/20%/20%.**
- [ ] **It would be reasonable to have a train/dev/test data split of roughly 90%/5%/5%.**
- [ ] **The test set should include some images from A and B.**
- [ ] **The training set should be made of images of A and B only.**

Clear selection

## ☆ Classification Metrics

You are designing a model that diagnoses whether a patient requires high risk surgery or not. Assume that:

- the binary label for a patient that needs the surgery is 1 (Positive)
- the binary label for a patient that could use a less aggressive medical treatment is 1 (Negative)

The single most important factor to take into account is that the model should not suggest the surgery option to a the patient that could receive alternative treatment.

What accuracy metric would you maximize? (TP means “True Positive”, FN means “False Negative”, FP means “False Positive”, TN means “True Negative”.)

Pick one of the choices

- TP/(TP+FP)
- TN/(TP+FN)
- TP/(TN+FP)
- TP/(TP+FN)

Clear selection

## ☆ Bayes error

Which of the following statements, if any, are true about the Bayes error? (Check all that apply.)

Pick the correct choices

- [ ] The test error of a ML algorithm can be smaller than the Bayes error.
- [x] The training error of a ML algorithm can be smaller than the Bayes error.
- [x] The validation error of a ML algorithm is usually larger than the Bayes Error.
- [ ] None of the above

Clear selection

Continue