Chapter 4 Pt. 1

Fundamentals of Machine Learning

Questions

1) The universal workflow/blueprint of machine learning. Write the correct letter for each step. (4 points)

Step 1. ___	a) Deciding on an evaluation protocol
Step 2. ___	b) Defining the problem and assembling a dataset
Step 3. ___	c) Developing a model that does better than a baseline
Step 4. ___	d) Choosing a measure of success
Step 5. ___	e) Preparing your Data
Step 6. ___	f) Regularizing your model and tuning your hyperparameters
Step 7. ___	g) Scaling up: developing a model that overfits

2) Speech recognition, language translation, optical character recognition, and image classification are all examples of what type of machine learning? (1 point)

   a. Unsupervised learning
   b. Supervised learning
   c. Self-supervised learning
   d. Reinforcement learning

3) Unsupervised learning consists of finding interesting transformations of the input data without the help of any _____. (1 point)

   a. Inputs
   b. Outputs
   c. Targets
   d. Loss function

4) Unsupervised learning is useful for what? (1 point)

   a. Data visualization
   b. Data compression
   c. Data denoising
   d. All the above
   e. None

5) Sequence generation, syntax tree prediction, object detection, and image segmentation are all examples of what type of machine learning? (1 point)

a. Unsupervised learning
b. Supervised learning
c. Self-supervised learning
d. Reinforcement learning

6) Match the following. (1 point)

____ Unsupervised learning	a) Classification Regression
____ Supervised learning	b) Dimensionality reduction Clustering

7) Match the terms and definitions. (6 points)

____ Sample or input	a) A task where the target is a set of continuous values.
____ Prediction or output	b) A classification task where each input sample should be categorized into more than two categories.
____ Target	c) A small set of samples (typically between 8 and 128) that are processed simultaneously by the model. The number of samples is often a power of 2, to facilitate memory allocation on GPU.
____ Prediction error or loss value	d) A specific instance of a class annotation in a classification problem.
____ Classes	e) A classification task where each input sample can be assigned multiple labels. The number of labels per image is usually variable.
____ Label	f) The truth. What your model should ideally have predicted, according to an external source of data.
____ Ground-truth or annotations	g) All targets for a dataset, typically collected by humans.
____ Binary classification	h) A set of possible labels to choose from in a classification problem.
____ Multiclass classification	i) What comes out of your model.
____ Multilabel classification	j) One data point that goes into your model.
____ Scalar regression	k) A task where the target is a continuous scalar value.
____ Vector regression	l) A classification task where each input sample should be categorized into two exclusive categories.
____ Mini-batch or batch	m) A measure of the distance between your model's prediction and the target.

Evaluating machine-learning models

8) In machine learning, the goal is to… (1 point)
   a. Achieve models that generalize
   b. Optimize performance on the training data
   c. Strictly follow the machine learning guidelines
   d. Measure the difference between outputs and inputs

9) What data sets should you have? (1 point)
   a. Training and test sets
   b. Training and validation sets
   c. Training, validation, and test sets
   d. Validation and test sets

10) What is the difference between hyperparameters and parameters? (1 point)
   a. Hyperparameters are the weights and parameters are the number of layers.
   b. Hyperparameters are the weights and parameters are the number and size of layers.
   c. Hyperparameters are the number and size of layers and parameters are the weights.
   d. Hyperparameters are the number of layers and parameters are the size of layers.

11) Tuning the configuration of the model based on its performance on the validation set can cause what problem(s)? (1 point)
   a. Underfitting
   b. Overfitting
   c. All the above
   d. None

12) What are information leaks? (1 point)
   a. Input data that is no longer being used in the model i.e. leaks
   b. Tuning hyperparameters causes information from the validation data to leak into the model
   c. Tuning hyperparameters causes information from the training data to leak into the model
   d. Tuning hyperparameters causes information from the test data to leak into the model

13) What are the three types of evaluation methods? (1 point)
   a. Simple hold-in validation, K-fold validation, and iterated K-fold validation with shuffling
   b. Complex hold-in validation, K-fold validation, and iterated K-fold validation with shuffling

c. Simple hold-out validation, K-fold validation, and randomized K-fold validation with shuffling
d. Simple hold-out validation, K-fold validation, and iterated K-fold validation with shuffling

14) Write the type of validation method used in the code example below. (1 point). Connect the code descriptions to the most appropriate code arrow or line. One description has been completed for you. (2 points)

_____

```
k = 4
num_validation_samples = len(data) // k

np.random.shuffle(data)

validation_scores = []
for fold in range(k):
    validation_data = data[num_validation_samples * fold:
     num_validation_samples * (fold + 1)]
    training_data = data[:num_validation_samples * fold] +
     data[num_validation_samples * (fold + 1):]

    model = get_model()
    model.train(training_data)
    validation_score = model.evaluate(validation_data)
    validation_scores.append(validation_score)


    validation_score = np.average(validation_scores)

    model = get_model()
    model.train(data)
    test_score = model.evaluate(test_data)
```

Creates a brand-new instance of the model (untrained)

Trains the final model on all non-test data available

Selects the validation-data partition

Validation score: average of the validation scores of the k folds

Uses the remainder of the data as training data. Note that the + operator is list concatenation, not summation.

15) Write the type of validation method used in the code example below. (1 point). Connect the code descriptions to the most appropriate code arrow or line. (2 points)

_____

```
num_validation_samples = 10000

np.random.shuffle(data)

validation_data = data[:num_validation_samples]
data = data[num_validation_samples:]

training_data = data[:]

model = get_model()
model.train(training_data)
validation_score = model.evaluate(validation_data)

# At this point you can tune your model,
# retrain it, evaluate it, tune it again...

model = get_model()
model.train(np.concatenate([training_data,
                            validation_data]))
test_score = model.evaluate(test_data)
```

Once you've tuned your hyperparameters, it's common to train your final model from scratch on all non-test data available.

Defines the training set

Trains a model on the training data, and evaluates it on the validation data

Shuffling the data is usually appropriate.

Defines the validation set