Chapter 4 Pt. 2

Fundamentals of Machine Learning

Questions

1) The universal workflow/blueprint of machine learning. Choose the characteristics that match each step. (6 points)

Step 1. Defining the problem and assembling a dataset

- Classification Accuracy.
- Logarithmic Loss.
- Area Under ROC Curve.
- Confusion Matrix.
- Classification Report.

Step 2. Choosing a measure of success

- Statistical power
- Last-layer activation
- Loss function
- Optimization configuration

Step 3. Deciding on an evaluation protocol

- Format data as tensors
- Scale data
- Homogenize data
- Feature normalization

Step 4. Preparing your data

- Simple hold-out validation set
- K-fold cross-validation
- Iterated K-fold validation

Step 5. Developing a model that does better than a baseline

- Add layers
- Make the layers bigger
- Train for more epochs

Step 6. Scaling up: developing a model that overfits

- Is there enough predictive data/info?
- Data is informative?
- Nonstationary?

- Add dropout
- Try different architectures
- Add L1 and/or L2
- Add/remove features
- Adjust units per layer or learning rate

Step 7. Regularizing your model and tuning your hyperparameters

2) Fill in the blanks for last-layer activation and loss function. (2 points)

Problem type	Last-layer activation	Loss function
Binary classification	`sigmoid`	_____
Multiclass, single-label classification	_____	_____
Multiclass, multilabel classification	_____	`binary_crossentropy`
Regression to arbitrary values	_____	_____
Regression to values between 0 and 1	_____	_____

3) Write the letter to connect related terms and protocols. (1 point)

____ Redundancy in your dat

____ The arrow of time

____ Data representativenes

a. Randomly shuffle data for multi-class classification.

b. Stock Future stock prices requires test data to be posterior to training data.

c. Make sure training and validation data are disjoint.

4) Data preprocessing includes… (1 point)
   a. Vectorization
   b. Normalization
   c. Handling missing values
   d. Feature extraction
   e. All the above

5) All inputs and targets in a neural network must be… (1 point)
   a. Tensors of floating-point data
   b. Tensors of floating-point data or integers
   c. Classified
   d. Clustered

6) The following code is an example of? (1 point)

```
x -= x.mean(axis=0)
x /= x.std(axis=0)
```

   a. Vectorization
   b. Normalization
   c. Randomization
   d. Featurization
   e. All the above

7) Homogenizing features, casting data to float32 with values of 0 to 1, and adjusting data to have a standard deviation of 1 and a mean of zero are all examples of? (1 point)
   a. Vectorization
   b. Normalization
   c. Randomization
   d. Featurization

e. All the above

8) How should you handle missing values in your data? (1 point)
   a. Input missing values into training data
   b. Allow the network to learn from exposure
   c. Let the network learn to ignore missing values
   d. Artificially generate training samples with missing entries
   e. All the above

9) What is feature engineering not? (1 point)
   a. The process of using your own knowledge about the data and about the machine-learning algorithm to make the algorithm work better
   b. Applying hardcoded (nonlearned) transformations to the data before it goes into the model.
   c. Presenting the data in a way that makes the model's job easier
   d. Making a problem easier by expressing it in a simpler way
   e. None

10) What is feature engineering? (1 point)
   a. Usually requires understanding the problem in depth
   b. Something deep learning removes the need for mostly
   c. Something neural networks automatically extract
   d. Can help solve a problem with far less data
   e. All the above

11) When the performance of a model on validation data peaks and starts to decline, this is called? (1 point)
   a. Generalization
   b. Overfitting
   c. Underfitting
   d. Regularization

12) Overfitting occurs… (1 point)
   a. Only in certain machine-learning problems
   b. In every machine-learning problem
   c. Before underfitting
   d. During data normalization

13) How are optimization and generalization different? (1 point)
   a. No difference
   b. Generalization is the process of adjusting a model to maximize performance on training data and optimization is the process of adjusting the model on validation data
   c. Generalization is the correlation between the rate of optimization and how well the model performs on data it has never seen before
   d. Generalization refers to how well the trained model performs on data it has never seen before and optimization is the process of adjusting the model to maximize performance on training data

14) Generalization is something… (1 point)
   a. The machine-learning engineer controls
   b. The machine-learning engineer cannot control

c. That causes underfitting
d. That depends on overfitting

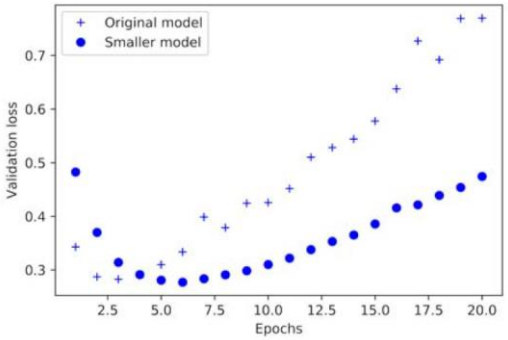
15) The best solution to prevent overfitting is to… (1 point)
a. Add in missing data points
b. Randomize weights
c. Get more training data
d. Create more layers

16) What is regularization? (1 point)
a. The process of fighting overfitting
b. The process of fighting underfitting
c. The process of normalization
d. The process of vectorizing

17) What is the simplest way to prevent overfitting? (1 point)
a. Create a new model
b. Reduce the size of the model
c. Add layers to the model
d. Adjust the weights of the model

18) What does capacity refer to in deep learning? (1 point)
a. The number of learnable parameters in a model
b. The number of weights in a model
c. The number of outputs in a model
d. The number of inputs in a model

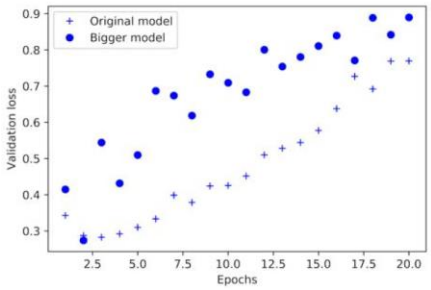
19) What is true about the figure below? (1 point)



a. The smaller model overfits more quickly.
b. The smaller model overfits more slowly.
c. The original model's performance increases over time.
d. The original model's data seems more reliable.

20) The following code and resulting figure is an example of? (1 point)

```
model = models.Sequential()
model.add(layers.Dense(512, activation='relu', input_shape=(10000,)))
model.add(layers.Dense(512, activation='relu'))
model.add(layers.Dense(1, activation='sigmoid'))
```



a. Generalization
b. Unreliable data
c. Too low capacity
d. Too much capacity

21) Which statement best summarizes the principle of Occam's razor? (1 point)
a. Details are better
b. Change is better
c. More is better
d. Simple is better

22) Weight regularization reduces overfitting by? (1 point)
a. Putting constraints on the complexity of a network
b. Forcing the network's weights to take only small values
c. Making the distribution of weight values more regular
d. Adding a cost to the loss function for larger weights
e. All the above

23) What is L1 regularization? (1 point)
a. Adding a cost proportional to the absolute value of the weight coefficients
b. Adding a cost proportional to the square of the value of the weight coefficients
c. Adding a weight decay to the neural network
d. All the above

24) What is L2 regularization? (1 point)
a. Adding a cost proportional to the absolute value of the weight coefficients
b. Adding a cost proportional to the square of the value of the weight coefficients
c. Adding a weight decay to the neural network
d. b & d
e. All the above

25) What is false about L1 and L2 regularization? (1 point)
a. They cannot be used in the same instance
b. The penalties they create are only added at training time
c. They cause an increased rate of underfitting
d. You must import 'regularizers' from keras

26) What is false about dropout? (1 point)
a. Randomly dropping out a number of output features during training

b. Also drop units during test time
c. The dropout rate is the fraction of the features that are zeroed out
d. The layer’s output values are scaled down by a factor equal to the dropout rate during test time

27) Using keras, where should dropout be implemented? (1 point)
a. After output of the layer and before the next
b. After the last layer
c. Before the first layers
d. All the above

28) What are the most common ways to prevent overfitting? (1 point)
a. Get more training data
b. Reduce the capacity of the network
c. Add weight regularization
d. Add dropout
e. a, c, & d
f. All the above
g. None