1. Select a correspondence of a task on the top to the definition of a task on the left:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Given a set of records consisting of several features and their categorical labels predict the labels for different set of records | Given an unlabeled set of records find its internal structure | Given a set of records labeled by a continuous variable predict labeling for an unknown set of records | Learn what set of actions to perform to maximize the reward | Given a set of records with any type of labels predict labels for other records |
| Supervised learning |  |  |  |  |  |
| Unsupervised learning |  |  |  |  |  |
| Regression |  |  |  |  |  |
| Classification |  |  |  |  |  |
| Reinforcement learning |  |  |  |  |  |

1. We have the following set of vectors and matrices:

x = [1, 2], y = [0, 1],

A = [[2, 0], [0, 1]]

Find the correspondence between the operations on these vectors/matrices on the left and the result at the top:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | [2, 2] | [[2], [2]] | [[0, 1], [0, 2]] | [[2, 0], [0, 1]] | Error | 2 | 1 |
| x·A |  |  |  |  |  |  |  |
| A·x |  |  |  |  |  |  |  |
| x·y |  |  |  |  |  |  |  |
| x·yT |  |  |  |  |  |  |  |
| xT·y |  |  |  |  |  |  |  |
| A·xT |  |  |  |  |  |  |  |
| x·A·yT |  |  |  |  |  |  |  |

1. Choose cost function for corresponding technique:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Linear Regression | Logistic Regression | Support Vector Machine | L1 regularization | L2 regularization |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

1. What each of the following learning curves mean?

|  |  |  |  |
| --- | --- | --- | --- |
|  | Something is wrong | Can be used to find the optimal # of coefficients | Can be used to find optimal regularization parameter (λ) |
| ncoef  J  training  Cross-validation |  |  |  |
| ncoef  J  Cross-validation  training |  |  |  |
| λ  J  training  Cross-validation |  |  |  |
| λ  J  training  Cross-validation |  |  |  |

1. What these plots are showing:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | No correlation between x and y | Possible outliers | Overfitting (high-variance) | Underfitting (high-bias) | Normal linear regression |
| x  y |  |  |  |  |  |
| x  y |  |  |  |  |  |
| x  y |  |  |  |  |  |
| x  y |  |  |  |  |  |
| x  y |  |  |  |  |  |

1. Which decision tree gives the best classification of the following data:

|  |  |  |  |
| --- | --- | --- | --- |
| Age | Sex | Weight | Like movies |
| 5 | M | 50 | Yes |
| 20 | M | 250 | Yes |
| 20 | F | 200 | Yes |
| 50 | M | 190 | No |
| 50 | F | 150 | No |

Sex?

Weight>195?

Age<30?

M

F

Yes

Yes

No

No

Sex?

Weight>195?

Age<30?

F

M

Yes

Yes

No

No

Sex?

Weight>195?

Age<30?

M

F

No

Yes

Yes

No

Sex?

Weight>195?

Age<30?

M

F

Yes

No

No

Yes

1. Which cluster dendrogram is the best representation of the following data:

x

y

1

2

3

4

5

6

1

6

4

5

2

3

1

3

4

5

6

2

1

2

5

4

6

3

1

2

4

5

6

3

4

6

1

5

2

3

1. Check where PCA may be necessary (check all that apply):

Plot data with many features

Forecast time series

Reduce the number of features for expensive neural network

Classify images

1. Check the best coefficients of the network allowing to correctly classify the following data:

x1

x2

1

2

1

2

1

x1

a1

h

w0

w2

x2

w1

w0 = 0, w1 = -1, w2 = -1

w0 = 2, w1 = 1, w2 = -1

w0 = 2, w1 = -1, w2 = -1

w0 = 0, w1 = -1, w2 = 1

1. What will be the size of the output image size after using convolutional kernel of size 2x2pixels followed by maxpooling layer of size 2x2 pixels on an image of size 2x2 pixels with stride 2 pixels and padding of 1 pixel:

2x2

1x1

3x3

4x4