CSC 412 Machine Learning and Knowledge Discovery

Exam I

Name: _____

Total: 30 + 2 points

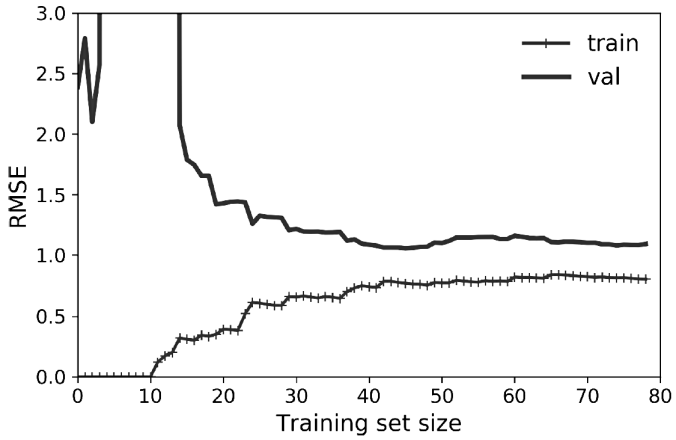
1. (1 point) What is the relationship between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL)?
   (a) AI $\supseteq$ ML $\supseteq$ DL
   (b) AI $\supseteq$ ML $\supset$ DL
   (c) AI $=$ ML $\supset$ DL
   (d) AI $\supset$ ML $\supset$ DL
   (e) AI $\supset$ ML $=$ DL

2. (1 point) What is supervised learning, and what are its typical learning tasks?

3. (1 point) How to do a 3-fold cross-validation?

4. (1 point) What will happen if a learning rate for Gradient Descent is set too high?

5. (1 point) What is the story behind the figure below?



6. (1 point) What does this talk about in the context of neural networks?

   "... If $f(x) = 2x + 3$ and $g(x) = 6x - 1$, then chaining these two functions gives you another function: $f\big(g(x)\big) = 2(6x - 1) + 3 = 12x + 1$. ..."

7. (1 point) **"We may want to reconsider the tradeoff between spending time and money on algorithm development versus spending it on corpus (data) development."** What does this comment mean?

8. (1 point) **"You must resist the temptation to tweak the hyperparameters to make the numbers look good on the test set."** Why?

9. (3 points) **Cross Entropy Cost Function**

Cross entropy is frequently used to measure how well a set of estimated class probabilities matches the target classes. Let the cross entropy cost function be

$$J(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} \boldsymbol{y}_k^{(i)} \cdot \log\left(\hat{\boldsymbol{p}}_k^{(i)}\right)$$

In this equation:

$\boldsymbol{y}_k^{(i)}$ is the target probability that the i-th instance belongs to class k. In general, it is either equal to 1 or 0, depending on whether the instance belongs to the class or not.

Here is a table of test results:

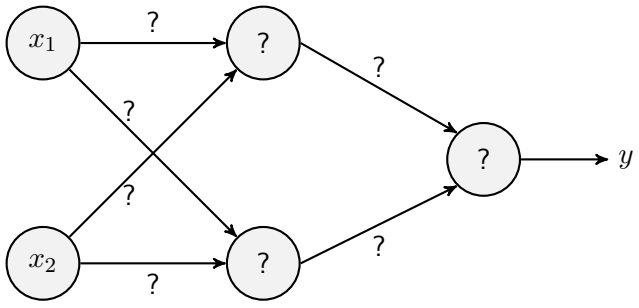
(a) What is m and What is K?

(b) What is the cost of this test result?

$\hat{\boldsymbol{p}}_1$	$\hat{\boldsymbol{p}}_2$	$\hat{\boldsymbol{p}}_3$	$\boldsymbol{y}_1$	$\boldsymbol{y}_2$	$\boldsymbol{y}_3$
0.2	0.6	0.2	0	1	0
0.4	0.3	0.3	0	0	1

10. (2 points) **Threshold Logic Units**

What are the weights and the thresholds of the neural network below:
(You can only use these numbers: -2, -1, 0, +1, and +2; Write your numbers beside the question marks)



(a)

x_1	x_2	y
0	0	1
0	1	0
1	0	0
1	1	1

(b)

11. (6 points) **Linear Regression**

After we trained a Linear Regression model, we got a parameter vector: $\hat{\boldsymbol{\theta}}^{\mathsf{T}} = [\, a, b, c \,]$

Now we have a data set:

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{y}$
d	e	f
g	h	i
j	k	l

(a) What is $\boldsymbol{x}_2^{(1)}$?

(b) What is $\hat{\boldsymbol{y}}^{(3)}$?

(c) List the new features if we transform the data to fit a Third-Degree Polynomial Regression.

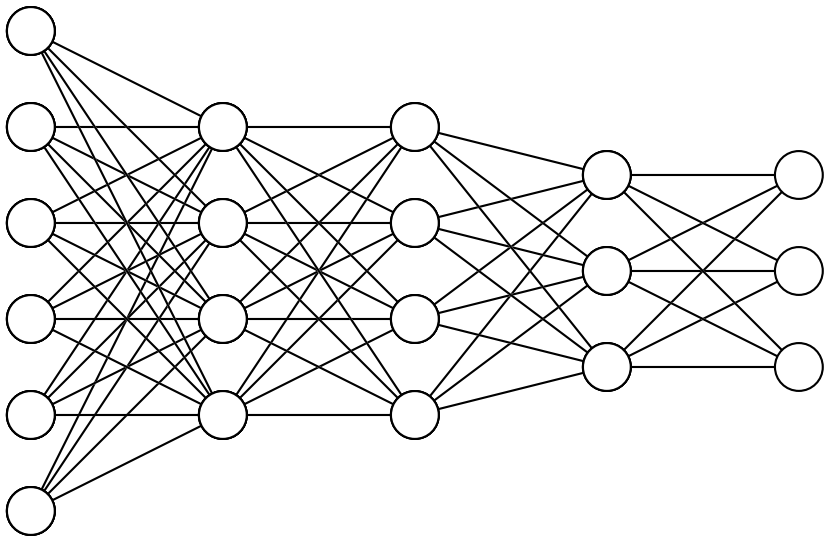
(d) We can perform Linear Regression by computing a closed-form equation:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}$$

If we regard this data set as a training set, how do we use this closed-form equation? (Just plug in the letters)

(e) What is the $\boldsymbol{\theta}^{(\text{next step})}$, if we do Batch Gradient Descent? Let the learning rate be α. (Hint: $\nabla_{\boldsymbol{\theta}} = \boldsymbol{X}^{\mathsf{T}}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})$; Just plug in the letters)

12. (9 points) **Deep Neural Network**



Here is a graph representing a fully connected network. All the hidden layers use the Sigmoid function, and the output layer uses the Softmax function. We use a data batch size of **1024** as an input.

(a) What is the shape of $\boldsymbol{X}$?

(b) What is the shape of $\boldsymbol{W}^{[3]}$?

(c) What is the shape of $\boldsymbol{A}^{[2]}$?

(d) What is the $\boldsymbol{a}_2^{[2]}$? (Darken the node)

(e) What are $\boldsymbol{W}_2^{[2]}$? (Darken the edges)

(f) How many parameters are there in this network?

(g) Complete the code of this network:

```
model = keras.models.Sequential([
    keras.layers.Flatten(input_shape=[2, 3]),
    keras.layers.Dense(4, activation="sigmoid"),
```

```
])
model.compile(loss="sparse_categorical_crossentropy",
                  optimizer="sgd", metrics=["accuracy"])
-----------------------------------------------------------
```

(h) Let $\boldsymbol{a}^{[3]} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\boldsymbol{W}^{[4]} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, and $\boldsymbol{b}^{[4]} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. What is $\boldsymbol{a}_2^{[4]}$?

13. (2 points) **HT Function**

Show that

$$\mathsf{HT}(z) = 2\sigma(2z) - 1 = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

14. (2 points) **(Bonus)** $\mathsf{HT}'(z) = ?$