1.

Which of the following data characteristics does NOT contribute to increased risk of overfitting?

- Noisy features ✓
- Too few data points ✓
- Rare feature values ✓
- Lying on a low-dimensional manifold

所有 natural data，根据 manifold hypothesis，都在对 high-dim manifold 的 lower-dim manifold 上

2.

What does the "manifold hypothesis" say?

- Naturally occurring data (images, audio, text, etc.) lies on a low-dimensional manifold within the high dimensional space in which it is encoded
- There are many types of data types where deep learning works well, e.g., images, audio, text, etc.
- Deep learning models form a low dimensional manifold in the space of all possible models
- There many different reasons why deep learning works well for images, audio, text, etc.

3.

high-dim manifold : 一 dim = 784

Consider the manifold of digit images that lies within the space of all 28x28 pixel images with pixel values in the integers 0 through 255. What does the "manifold hypothesis" say in this particular case?

so the dim of this subspace: < 784

- The dimension of the manifold of digit images will be much smaller compared to 784
- The dimension of the manifold of digit images will be much larger compared to 784
- The dimension of the manifold of digit images will be much smaller compared to 255
- The dimension of the manifold of digit images will be much larger compared to 255

4.

Let A and B be two 28 x 28 NumPy arrays representing the digits 2 and 6 respectively. Which NumPy expression will compute the manifold interpolation between A and B?

- `(2 + 6) / 2`
- `(A + B) / 2`
- `np.sqrt(A * B)`
- There is no simple NumPy expression for it

continuous mapping ... (没有 simple expression)

2 . . . . . 6

5.

How many times do you have to train deep learning models when doing 4-fold cross-validation?

- 4
- 3
- 1
- 16

| | | | | |
|---|---|---|---|---|
| 1 | V | T | T | T |
| 2 | T | V | T | T |
| 3 | T | T | V | T |
| 4 | T | T | T | V |

V: validation T: training

6

Which of the following will NOT help with overfitting?

- Increasing model size
- Decreasing model size
- Early stopping
- Increasing training data ✓

减小模型复杂度，可防止 overfitting

↑ training dataset 可增进 model generalization 防止被投喂数据影响.

7.

In the MNIST experiment involving addition of 784 extra features, we made two choices: zero features and random noise features. Which choice led to more overfitting?

- zero features
- random noise features
- both overfit to the same extent
- there was no clear evidence either way

学习不到东西

会学习错误的 pattern.

8.

Which of the following is a key parameter related to minibatch stochastic gradient descent that has a direct impact on overfitting?

- number of epochs
- number of layers
- sizes of each layer
- type of problem (classification versus regression)

epoch 过多 → 单T epoch 的大小过小 → 更容易 learn from noises.

9. (根本不 iterate)

What happens, as training epochs proceed, if you choose a learning rate of zero?

- (●) Training loss will not go down
- ( ) Training loss will oscillate
- ( ) Validation loss will oscillate
- ( ) Validation loss will go up very quickly

10.

Suppose a deep learning model had the last layer like this:

```
layers.Dense(1)
```

no activation function : regression

What kind of problem is this network trying to solve?

- (●) Regression problem
- ( ) Binary classification problem
- ( ) The code will yield an error because activation function was not specified
- ( ) Insufficient information to arrive at an answer