1. If you have 10,000 examples, how would you split the train/dev/test set? Choose the best option. **1 / 1 point**

- 33% train. 33% dev. 33% test.
- 60% train. 20% dev. 20% test.
- 98% train. 1% dev. 1% test.

Expand

Correct
Yes. This might be considered a small data set, not in the range of big data. Thus a more classical (old) best practice should be used.

2. In a personal experiment, an M.L. student decides to not use a test set, only train-dev sets. In this case which of the following is true? **1 / 1 point**

- He won't be able to measure the variance of the model.
- He might be overfitting to the dev set.
- He won't be able to measure the bias of the model.
- Not having a test set is unacceptable under any circumstance.

Expand

Correct
Yes. Although not recommended, if a more accurate measure of the performance is not necessary it is ok to not use a test set. However, this might cause an overfit to the dev set.

3. A model developed for a project is presenting high bias. One of the sponsors of the project offers some resources that might help reduce the bias. Which of the following additional resources has a better chance to help reduce the bias? **0 / 1 point**

- Use different sources to gather data and better test the model.
- Give access to more computational resources like GPUs.
- Gather more data for the project.

Expand

Incorrect
No. More data won't reduce the bias.

4. Working on a model to classify bananas and oranges your classifier gets a training set error of 0.1% and a dev set error of 11%. Which of the following two are true? **1 / 1 point**

- [x] The model is overfitting the train set.
  - **Correct**
    Yes. This is precisely what happens when overfitting.
- [x] The model has a high variance.
  - **Correct**
    No. This model has a low bias and high variance.
- [ ] The model is overfitting the dev set.
- [ ] The model has a very high bias.

Expand

Correct
Great, you got all the right answers.

5. Which of the following are regularization techniques? **1 / 1 point**

- [x] Weight decay.
  - **Correct**
    Correct. Weight decay is a form of regularization.
- [ ] Gradient Checking.
- [x] Dropout.
  - **Correct**
    Correct. Using dropout layers is a regularization technique.
- [ ] Increase the number of layers of the network.

Expand

Correct
Great, you got all the right answers.

6. What happens when you increase the regularization hyperparameter lambda? **1 / 1 point**

- Weights are pushed toward becoming smaller (closer to 0)
- Gradient descent taking bigger steps with each iteration (proportional to lambda)
- Doubling lambda should roughly result in doubling the weights
- Weights are pushed toward becoming bigger (further from 0)

Expand

Correct

7. With the inverted dropout technique, at test time: **0 / 1 point**

- You do not apply dropout (do not randomly eliminate units), but keep the 1/keep_prob factor in the calculations used in training.
- You apply dropout (randomly eliminating units) and do not keep the 1/keep_prob factor in the calculations used in training
- You apply dropout (randomly eliminating units) but keep the 1/keep_prob factor in the calculations used in training.
- You do not apply dropout (do not randomly eliminate units) and do not keep the 1/keep_prob factor in the calculations used in training

Expand

Incorrect

8. Increasing the parameter keep_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply) **1 / 1 point**

- [ ] Increasing the regularization effect
- [x] Reducing the regularization effect
  - **Correct**
- [ ] Causing the neural network to end up with a higher training set error
- [x] Causing the neural network to end up with a lower training set error
  - **Correct**

Expand

Correct
Great, you got all the right answers.

9. Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) **1 / 1 point**

- [x] Data augmentation
  - **Correct**
- [ ] Xavier initialization
- [ ] Vanishing gradient
- [ ] Exploding gradient
- [x] Dropout
  - **Correct**
- [x] L2 regularization
  - **Correct**
- [ ] Gradient Checking

Expand

Correct
Great, you got all the right answers.

10. Which of the following is the correct expression to normalize the input **X**? **1 / 1 point**

- $x = \frac{x}{\sigma}$
- $x = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$
- $x = \frac{x - \mu}{\sigma}$
- $x = \frac{1}{m}\sum_{i=1}^{m} (x^{(i)})^2$

Expand

Correct
Correct. This shifts the mean of the input to the origin and makes the variance one in each coordinate of the input examples.