Q1-1: Are these statements true or false?
(A) We can add noise to the input to get a solution of higher confidence among multiple optimal solutions.
(B) More noise added to the input will lead to a more robust solution.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false?
(A) We can add noise to the input to get a solution of higher confidence among multiple optimal solutions.
(B) More noise added to the input will lead to a more robust solution.

1. True, True
2. True, False 
3. False, True
4. False, False

(A) As is shown in the lecture.
(B) Too much noise may cause data cross the boundary and make it hard to classify.

Q1-2: Are these statements true or false?
For the hypothesis $f = w^T x$,
(A) Adding Gaussian noise $\varepsilon \sim \mathcal{N}(0, \eta I)$ to the input is equivalent to weight decay regularization.
(B) Adding Gaussian noise $\varepsilon \sim \mathcal{N}(0, \eta I)$ to the weight is equivalent to adding a regularization term on the norm of w.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-2: Are these statements true or false?
For the hypothesis $f = w^T x$,
(A) Adding Gaussian noise $\varepsilon \sim \mathcal{N}(0, \eta I)$ to the input is equivalent to weight decay regularization.
(B) Adding Gaussian noise $\varepsilon \sim \mathcal{N}(0, \eta I)$ to the weight is equivalent to adding a regularization term on the norm of w.

1. True, True
2. True, False ⟵
3. False, True
4. False, False

(A) As is derived in the lecture.
(B) By the derivation in the lecture, the regularization term is $\mathbb{E} \| \nabla f_w(x) \|^2$

Q2-1: Are these statements true or false?
(A) Adding noise to the input is also a kind of data augmentation.
(B) Data augmentation can help enrich limited training data, so can always improve the model performance.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-1: Are these statements true or false?

(A) Adding noise to the input is also a kind of data augmentation.
(B) Data augmentation can help enrich limited training data, so can always improve the model performance.

1. True, True
2. True, False ⬅
3. False, True
4. False, False

(A) As is shown in the lecture.
(B) It's true that data augmentation can help enrich limited training data, as it can simulate much more new data based on current data. But we need to be careful with the transformation, such rotate "6" to "9", otherwise it may introduce unwanted noise to influence the classification .

Q2-2: Are these statements true or false?
(A) We need validation data to decide when to early stop.
(B) We can think early stopping as a regularization to limit the volume of parameter space reachable from the initial parameter.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-2: Are these statements true or false?
(A) We need validation data to decide when to early stop.
(B) We can think early stopping as a regularization to limit the volume of parameter space reachable from the initial parameter.

1. True, True 
2. True, False
3. False, True
4. False, False

(A) As is shown in the lecture.
(B) That's true. Early stopping will limit the training time and thus potentially limit the space the training can search.

Q3-1: Are these statements true or false?
(A) In dropout, we can only mask the hidden units of the neural network.
(B) In batch normalization, we can drop the final step of shift and scaling. Because the aim of BN is just to reduce the internal covariate shift, and it has already been done by normalization of inputs of the unit.

1. True, True
2. True, False
3. False, True
4. False, False

Q3-1: Are these statements true or false?
(A) In dropout, we can only mask the hidden units of the neural network.
(B) In batch normalization, we can drop the final step of shift and scaling. Because the aim of BN is just to reduce the internal covariate shift, and it has already been done by normalization of inputs of the unit.

1. True, True
2. True, False
3. False, True
4. False, False ⬅

(A) In dropout, we can mask both input units and hidden units.
(B) The step of shift and scaling is necessary, otherwise the normalization may make the layer lose the expression power. We can leave it to later layers themselves to choose whether to undo this step.

Q3-2: Are these statements true or false?
(A) For feature selection, we usually tend to use l_2 regularization.
(B) For object detection, data augmentation can be very useful to simulate data.

1. True, True
2. True, False
3. False, True
4. False, False

Q3-2: Are these statements true or false?
(A) For feature selection, we usually tend to use l_2 regularization.
(B) For object detection, data augmentation can be very useful to simulate data.

1. True, True
2. True, False
3. False, True 
4. False, False

(A) We usually use l_1 regularization for feature selection, as we want a sparse solution to see which features can be discarded.
(B) We can use data augmentation to easily simulate enormous variety of factors of variation, thus improving the model performance.