Question 1 — 10 pts

Often autoencoders are used for dimensionality reduction or feature learning. They can be viewed as a special case of feedforward networks and may be trained with all the same techniques, typically minibatch gradient descent following gradients computed by back-propagation. One way to obtain useful features from the autoencoder is to constrain the code h to have a larger dimension than the input x.

- ○ True
- ◉ False

Question 2 — 10 pts

In VAE, we define a lower bound on the log likelihood. This is a function that is always less than or equal to the log likelihood for a given value of ϕ and will also depend on some other parameters θ. Eventually, we will build a network to compute this lower bound and optimize it. To define this lower bound, we need Jensen's inequality. Which of the statement below is correct?

- ◉ A concave function g[•] of the expectation of the data y is greater than or equal to the expectation of the function of the data
- ○ A concave function g[•] of the expectation of the data y is less than or equal to the expectation of the function of the data

Question 3 — 10 pts

The KL divergence is a measure of "distance" between distributions and can only take non-negative values.

- ◉ True
- ○ False

Question 4 — 10 pts

If there are D inputs x• and D hidden units h•, then this fully connected layer would have how many weights ω•• and how many biases biases β•?

- ○ D weights, D biases
- ○ D weights, 1 bias
- ◉ D^2 weights, D biases
- ○ 2D weights, D biases

Question 5 — 10 pts

Following question 4, how many weights and biases do we have in the convolutional layer that has a kernel of size 4?

- ◉ 4 weights, 1 bias
- ○ 4D weights, D biases
- ○ D weights, D biases
- ○ D^2 weights, 1 bias

Question 6 — 10 pts

In stochastic gradient descent, at each iteration, the algorithm chooses a random subset of the training data and computes the gradient from these examples alone.

- ◉ True
- ○ False

Question 7

10 pts

A convex function g[•] of the expectation of the data y is greater than or equal to the expectation of the function of the data.

- ✗ True
- ● False

Question 8

10 pts

At saddle points, the gradient is zero, but the function increases in some directions and decreases in others.

- ◉ True
- ○ False

Question 9

10 pts

In CNN, zero padding assumes that the input is zero outside its valid range.

- ◉ True
- ○ False

Question 10

10 pts

Logarithm is convex.

- ○ True
- ◉ False