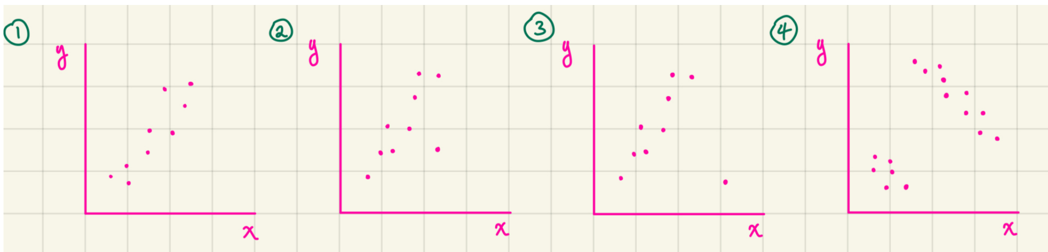
CS 1810 Concept Checks

1 Linear Regression

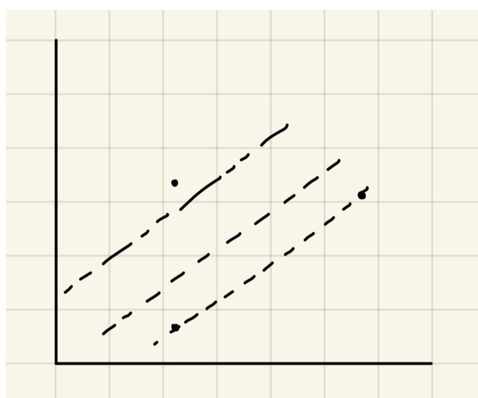
1. What would the linear regression model look like for each of the following datasets?



2. Consider the following distributions for the noise in a probablistic linear regression model:

   - **Laplace** ($\lambda = 1$): $p(\epsilon) \propto \exp(-|\epsilon|)$
   - **Gaussian** ($\sigma^2 = 1$): $p(\epsilon) \propto \exp(-\frac{1}{2}\epsilon^2)$
   - **Student-t** ($\nu = 1$): $p(\epsilon) \propto (1 + \epsilon^2)^{-1}$

   For each of the dotted lines below, explain which distribution would correspond with that particular model.



2 Classification

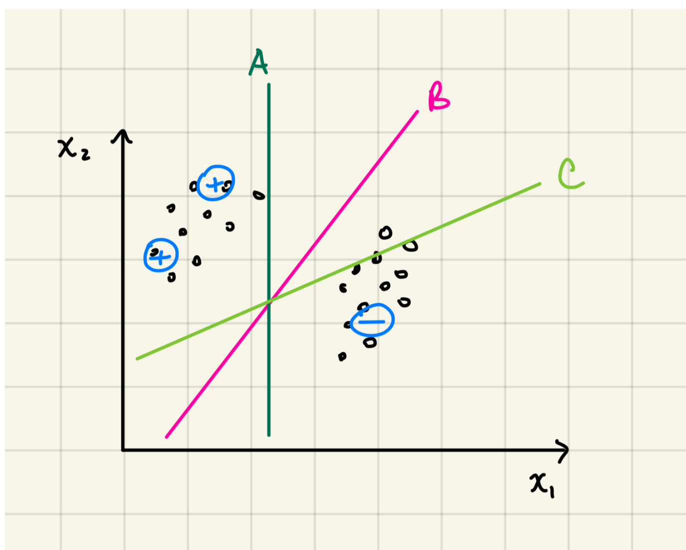
3. Fill out the following table to describe the different types of classification loss functions.

	$\ell_{0/1}$	ℓ_{hinge}	ℓ_{logic}
convex			
NP-hard			
differentiable			
cares about how "wrong"			
cares about how "right"			

4. Consider the following semi-supervised data set, meaning some of the points are labeled and some of them are not. We have three possible decision boundaries, lines A, B, and C. All of the boundaries are correct given labeled data.

   (a) Does seeing unlabeled data effect your preference?

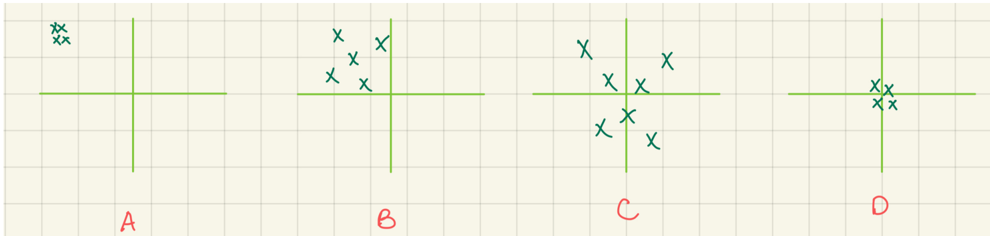
   (b) Does it matter/justify with respect to discriminative story vs. generative story?



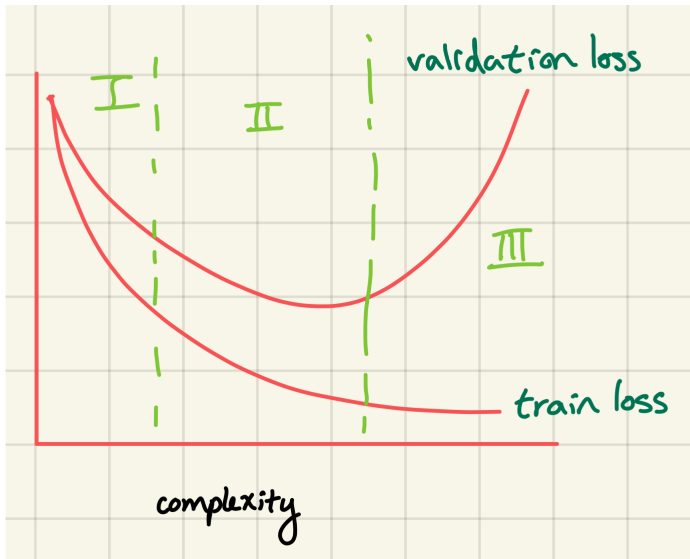
3 Model Selection and Neural Networks

5. Two questions:

   (a) For each of the four graphs below, our goal is the center of the graph and the green X's represent our prediction. Describe whether each graph has 1) high or low bias and 2) high or low variance.



(b) Label each out of I, II, III as overfitting, underfitting, or good fit.



6. Suppose we have two data points $(0,0),(1,1)$. Now consider the following 3 models:

- Model 1: $\hat{y} = a_0$
- Model 2: $\hat{y} = a_0 + a_1 x$
- Model 3: $\hat{y} = a_0 + a_1 x + a_2 x^2$

We also have the priors

$$p(a_0) = \text{Unif}(-100, 100)$$
$$p(a_1) = \text{Unif}(-1, 1)$$
$$p(a_2) = \text{Unif}(-100, 100)$$

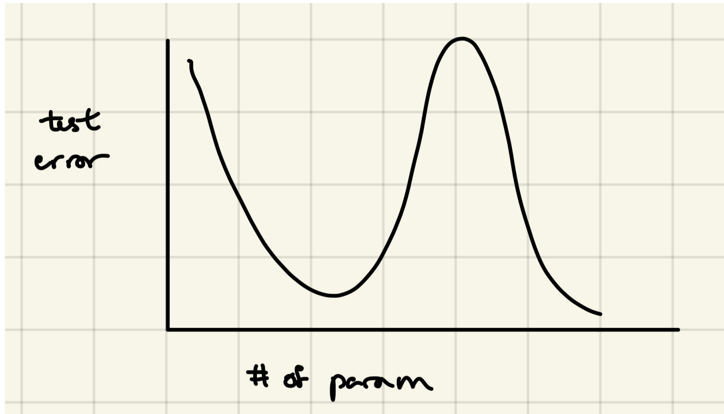
What is the posterior predictive at $x = 1/2$?

7. A couple of questions on bias-variance for neural networks:

(a) As the number of parameters increases, can bias increase?

(b) If the number of parameters is way less than the number of data points, can we fit the data perfectly? If the number of parameters is equal to the number of data points, can we fit the data perfectly? If the number of parameters is way greater than the number of data points, can we fit the data perfectly?

(c) Examine the following graph:



Why might we see this second descent in the loss?

4 Support Vector Machines

8. We are considering the regularization parameter C in the soft margin SVM formulation. For each of the following questions, answer with small C or big C:

   (a) Which would you choose if the goal is to maximize the margin?

   (b) Which would you choose if the goal is to minimize error on the training set?

   (c) Which would you choose if the goal is to overfit?

   (d) Which would you choose if the goal is to underfit?

   (e) Which would you choose if the goal is to have a less "flat" boundary?

   (f) Which would you choose if the goal is to have a more "flat" boundary?

9. A couple of questions on overfitting in SVMs:

   (a) What would happen if we use the RBF kernel with a small σ^2? (in terms of over and underfitting)

   (b) What would happen if we use the RBF kernel with a large σ^2? (in terms of over and underfitting)

   (c) How do we know if an SVM is overfitting?