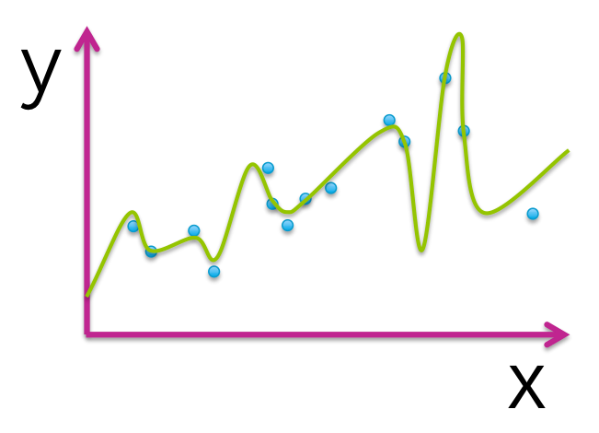
## Week 2 - Regression

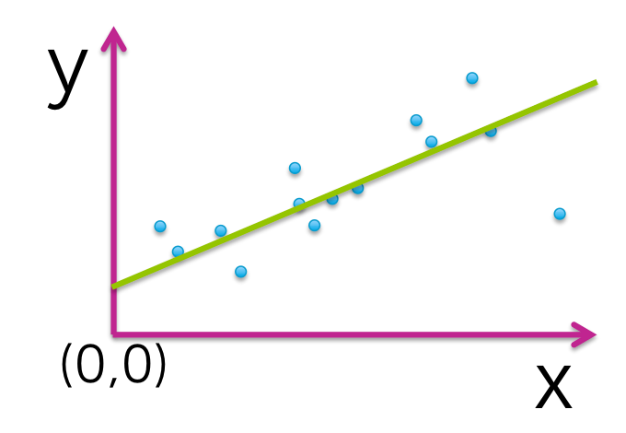
1. Which figure represents an overfitted model?



1. True or false: The model that best minimizes training error is the one that will perform best for the task of prediction on new data.
   1. False
2. The following table illustrates the results of evaluating 4 models with different parameter choices on some data set. Which of the following models fits this data the best?

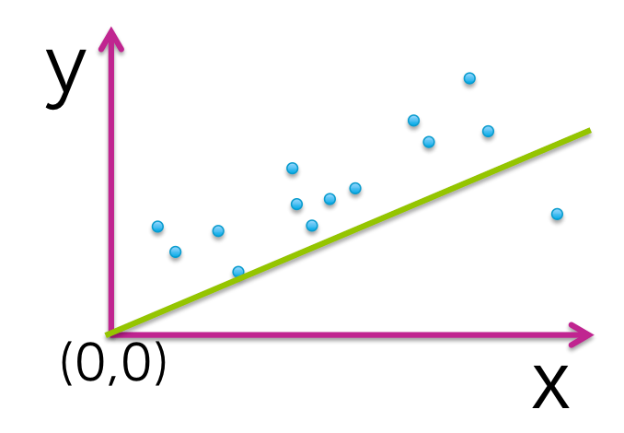
|  |  |  |
| --- | --- | --- |
| Model index | Parameters (intercept, slope) | Residual sum of squares (RSS) |
| 1 | 0, 1.4 | 20.51 |
| 2 | 3.1, 1.4 | 15.23 |
| 3 | 2.7, 1.9 | 13.67 |
| 4 | 0, 2.3 | 18.99 |

1. Model 3
2. Assume we fit the following quadratic function: f(x) = w0+w1\*x+w2\*(x^2) to the dataset shown (blue circles). The fitted function is shown by the green curve in the picture below. Out of the 3 parameters of the fitted function (w0, w1, w2), which ones are estimated to be 0?



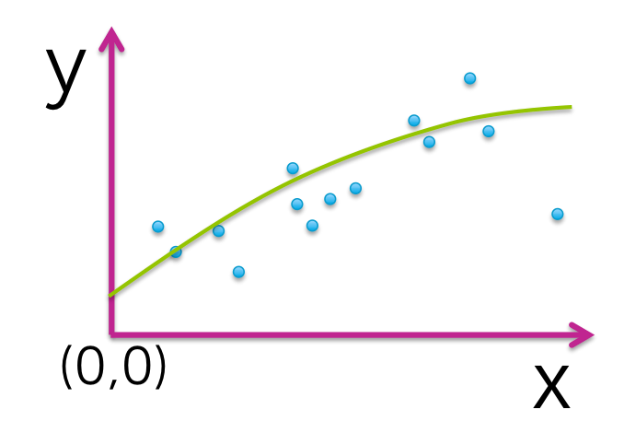
* 1. w2

1. Assume we fit the following quadratic function: f(x) = w0+w1\*x+w2\*(x^2) to the dataset shown (blue circles). The fitted function is shown by the green curve in the picture below. Out of the 3 parameters of the fitted function (w0, w1, w2), which ones are estimated to be 0?



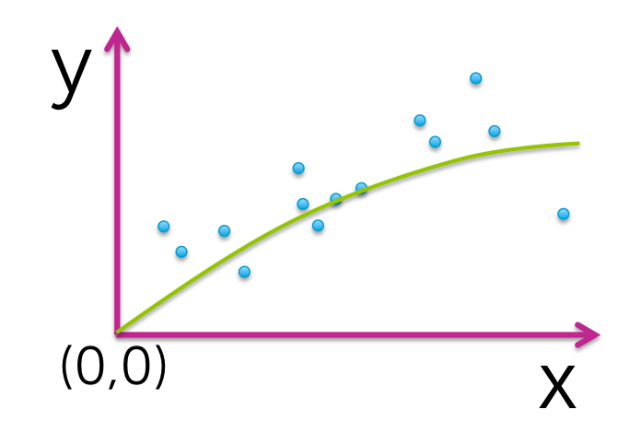
* 1. w0
  2. w2

1. Assume we fit the following quadratic function: f(x) = w0+w1\*x+w2\*(x^2) to the dataset shown (blue circles). The fitted function is shown by the green curve in the picture below. Out of the 3 parameters of the fitted function (w0, w1, w2), which ones are estimated to be 0?

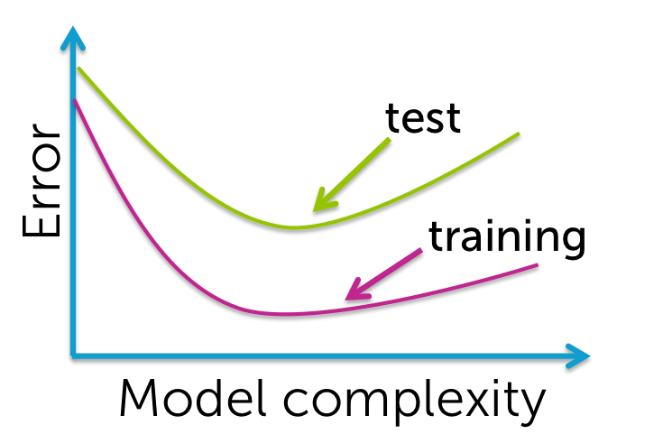


* 1. None of them

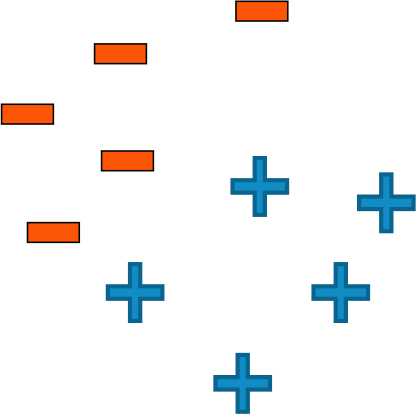
1. Assume we fit the following quadratic function: f(x) = w0+w1\*x+w2\*(x^2) to the dataset shown (blue circles). The fitted function is shown by the green curve in the picture below. Out of the 3 parameters of the fitted function (w0, w1, w2), which ones are estimated to be 0?



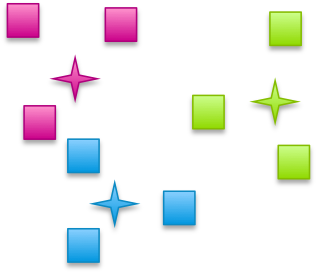
* 1. w0

1. Which of the following plots would you not expect to see as a plot of training and test error curves?
   1. 
2. True or false: One always prefers to use a model with more features since it better captures the true underlying process.
   1. False

## Week 3 – Classification

1. The simple threshold classifier for sentiment analysis described in the video:
   1. Must have pre-defined positive and negative attributes.
   2. Must either count attributes equally or pre-define weights on attributes.
2. For a linear classifier classifying between “positive” and “negative” sentiment in a review “x”, implies:
   1. We are uncertain whether the review is “positive” or “negative”.
3. For which of the following datasets would a linear classifier perform perfectly?
   1. 
4. High classification accuracy always indicates a good classifier. True/False?
   1. False.
5. For a classifier classifying between 5 classes, there always exists a classifier with accuracy greater than 0.18. True/False?
   1. True.
6. A false negative is always worse than a false positive. True/False?
   1. False.
7. Which of the following statements are true?
   1. Test error tends to decrease with more training data until a point, and then does not change (i.e. the curve flattens out).

## Week 4 – Clustering and Similarity

1. A country, called Simpleland, has a language with a small vocabulary of just “the”, “on”, “and”, “go”, “round”, “bus”, and “wheels”. For a word count vector with indices ordered as the words appear above, what is the word count vector for a document that simply says “the wheels on the bus go round and round.”
   1. 2 1 1 1 2 1 1
2. In Simpleland, a reader is enjoying a document with a representation: [1 3 2 1 2 1 1]. Which of the following articles would you recommend to this reader next?
   1. 1 7 0 0 2 0 1. (It has the highest element-wise product sum).
3. A corpus in Simpleland has 99 articles. If you pick one article and perform 1-nearest neighbour search to find the closest article to this query article, how many times must you compute the similarity between two articles?
   1. 98
4. For the TF-IDF representation, does the relative importance of words in a document depend on the base of the logarithm used? For example, take the words "bus" and "wheels" in a particular document. Is the ratio between the TF-IDF values for "bus" and "wheels" different when computed using log base 2 versus log base 10?
   1. No
5. Which of the following statements are true?
   1. Deciding whether an email is spam or not spam using the text of the email and some spam / not spam labels is a supervised learning problem.
   2. If we are performing clustering, we typically assume we either do not have or do not use class labels in training the model.
6. Which of the following pictures represents the best k-means solution? (Squares represent observations, plus signs are cluster centres, and colours indicate assignments of observations to cluster centres.)
   1. 

## Week 5 – Recommender Systems

1. Recommending items based on global popularity can:
   1. None of the above.
2. Recommending items using a classification approach can:
   1. Provide personalisation.
   2. Capture context (e.g. time of day).
3. Recommending items using a simple count-based co-occurrence matrix can:
   1. Provide personalisation.
4. Recommending items using featurised matrix factorisation can:
   1. Provide personalisation.
   2. Capture context (e.g. time of day).
5. Normalising co-occurrence matrices is used primarily to account for:
   1. Items purchased by many people.
6. A store has 3 customers and 3 products. Below are the learned feature vectors for each user and product. Based on the estimated model, which product would you recommend first to User #2?

|  |  |
| --- | --- |
| **User ID** | **Feature Vector** |
| 1 | (1.73, 0.01, 5.22) |
| 2 | (0.03, 4.41, 2.05) |
| 3 | (1.13, 0.89, 3.76) |

|  |  |
| --- | --- |
| **Product ID** | **Feature Vector** |
| 1 | (3.29, 3.44, 3.67) |
| 2 | (0.82, 9.71, 3.88) |
| 3 | (8.34, 1.72, 0.02) |

* 1. Product 2

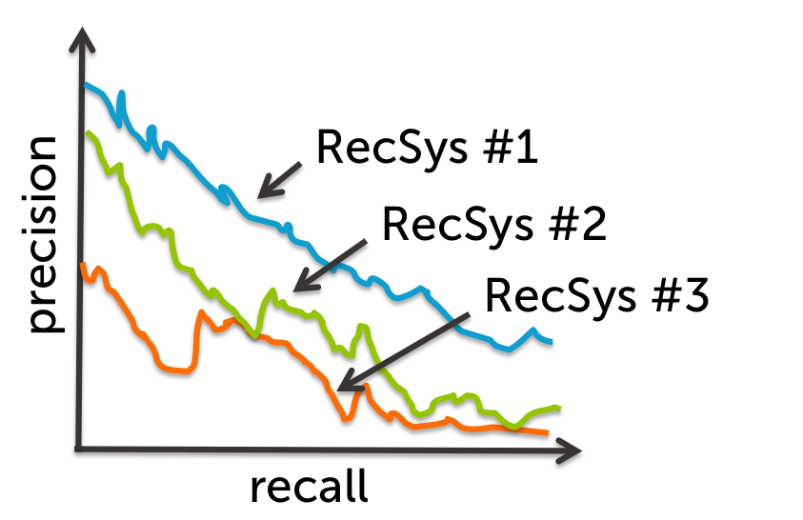
1. For the liked (magenta) and recommended (green) items displayed below, calculate the recall. (Items not recommended are greyed out for clarity).



1. For the liked (magenta) and recommended (green) items displayed below, calculate the precision. (Items not recommended are greyed out for clarity).



1. Based on the precision-recall curves in the figure, which recommender system would you use?



* 1. RecSys #1

## Week 6 – Deep Learning

1. Which of the following statements are true?
   1. Having good non-linear features can allow us to learn very accurate linear classifiers.
2. A linear classifier can represent which of the following functions?
   1. x1 OR x2 OR NOT x3
   2. x1 AND x2 AND NOT x3
   3. x1 OR (x2 AND NOT x3) – NOTE – this is actually incorrect.

Truth tables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X1 | X2 | X3 |  | x1 OR x2 OR NOT x3 | x1 AND x2 AND NOT x3 | x1 OR (x2 AND NOT x3) |
| 0 | 0 | 0 |  | 1 | 0 | 0 |
| 0 | 0 | 1 |  | 0 | 0 | 0 |
| 0 | 1 | 0 |  | 1 | 0 | 1 |
| 0 | 1 | 1 |  | 1 | 0 | 0 |
| 1 | 0 | 0 |  | 1 | 0 | 1 |
| 1 | 0 | 1 |  | 1 | 0 | 1 |
| 1 | 1 | 0 |  | 1 | 1 | 1 |
| 1 | 1 | 1 |  | 1 | 0 | 1 |

Diagrams

X3

(0, 0, 1)

(1, 1, 1)

(1, 1, 0)

(1, 0, 1)

(0, 1, 1)

(1, 0, 0)

(0, 1, 0)

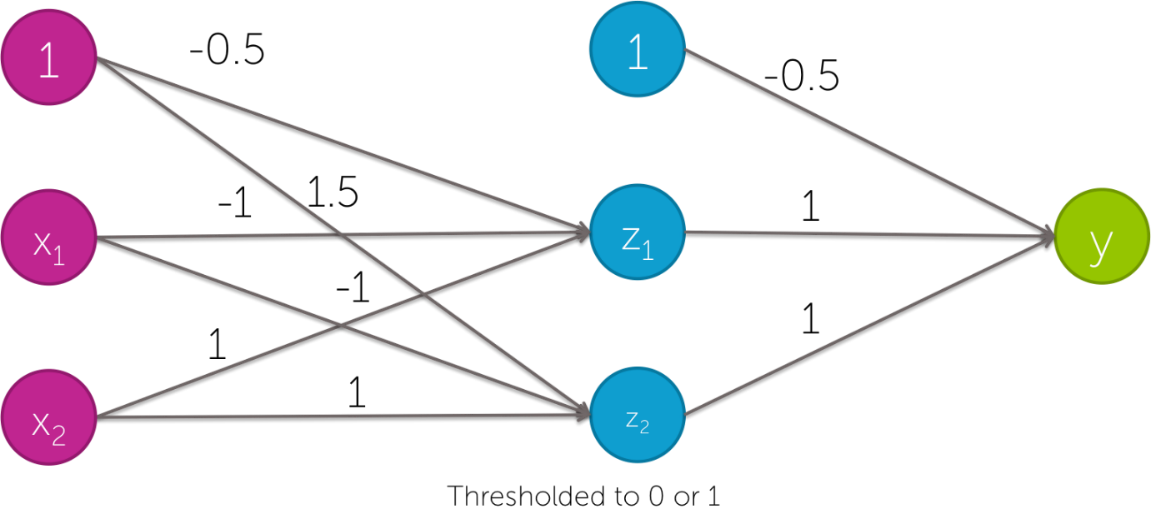
(0, 0, 0)

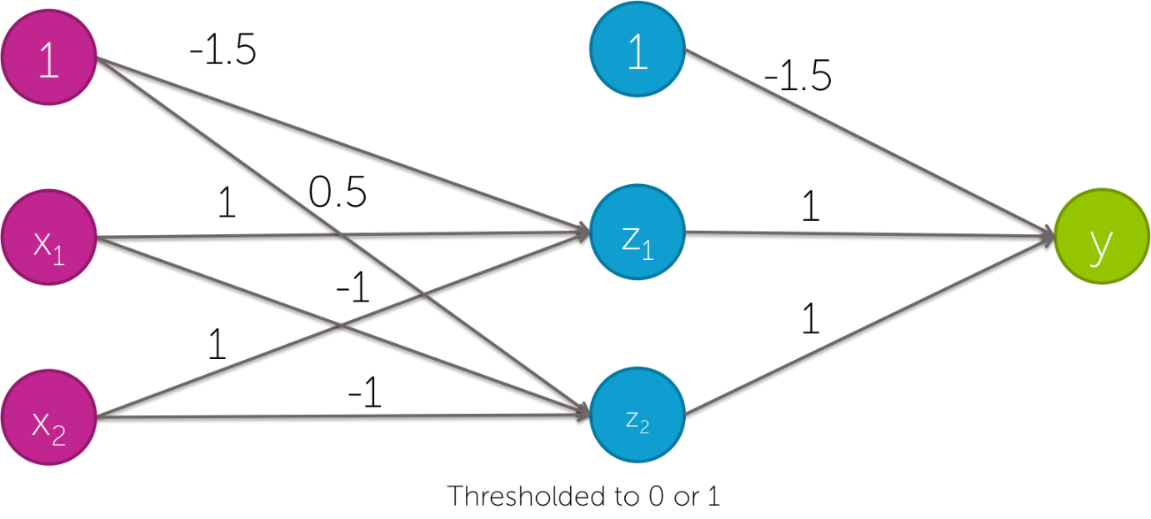
X2

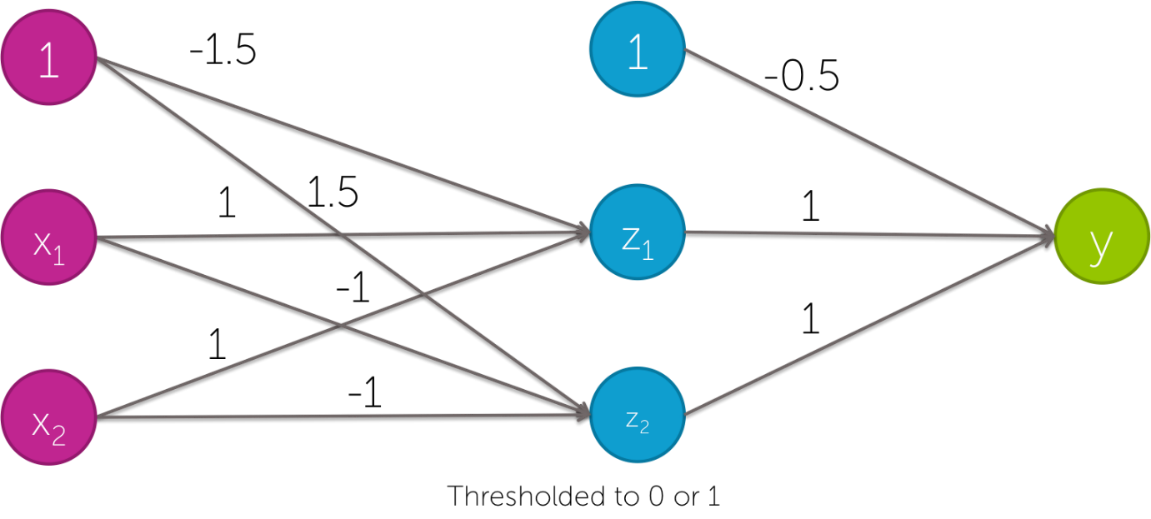
X1

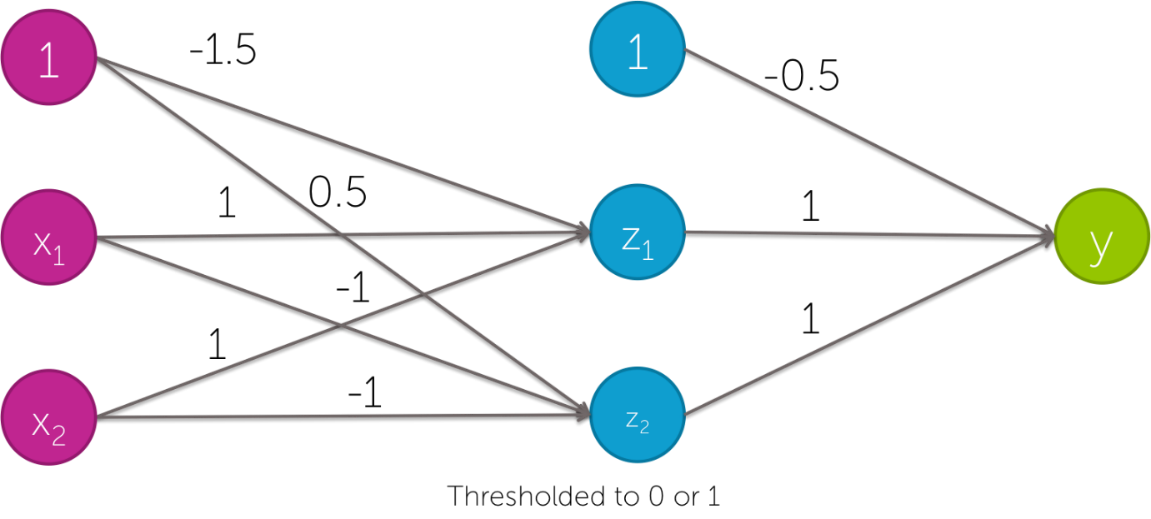
Does a plane exist that can separate the zeros from the ones for each function?

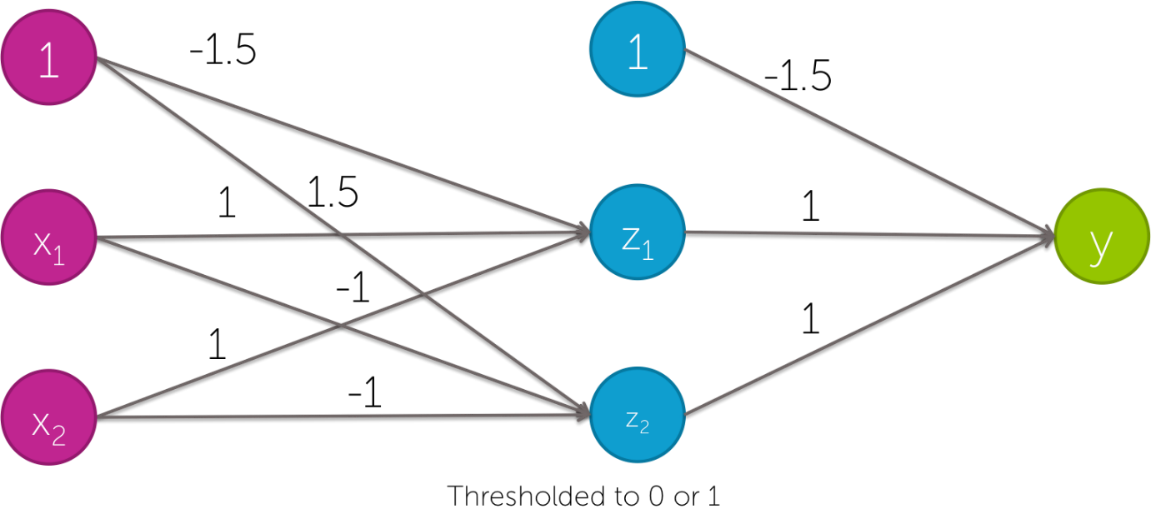
1. Which of the following neural networks can represent the function (x1 AND x2) OR (NOT x1 AND NOT x2)?











(x1 AND x2) OR (NOT x1 AND NOT x2) = z1 or z2 where

z1 = x1 AND x2  
 z2 = NOT x1 AND NOT x2

3D only, but still don’t understand the logic of why.

1. Which of the following statements are true?
   1. Features in computer vision act like local detectors.
   2. By learning non-linear features, neural networks have allowed us to automatically learn detectors for computer vision.
2. If you have lots of images of different types of plankton labelled with their species name, and lots of computational resources, what would you expect to perform better predictions:
   1. A deep neural network trained on this data.
3. If you have a few images of different types of plankton labelled with their species name, what would you expect to perform better predictions:
   1. A simple classifier trained on this data, using deep features as input that were trained on ImageNet data.