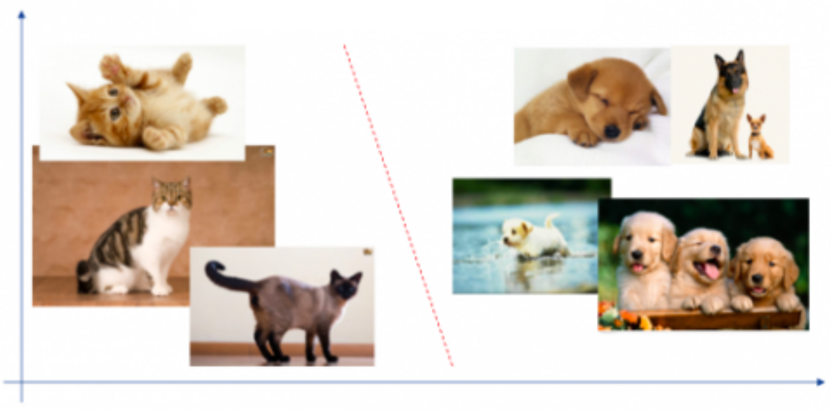
Machine Learning

Supervised Learning
Classification

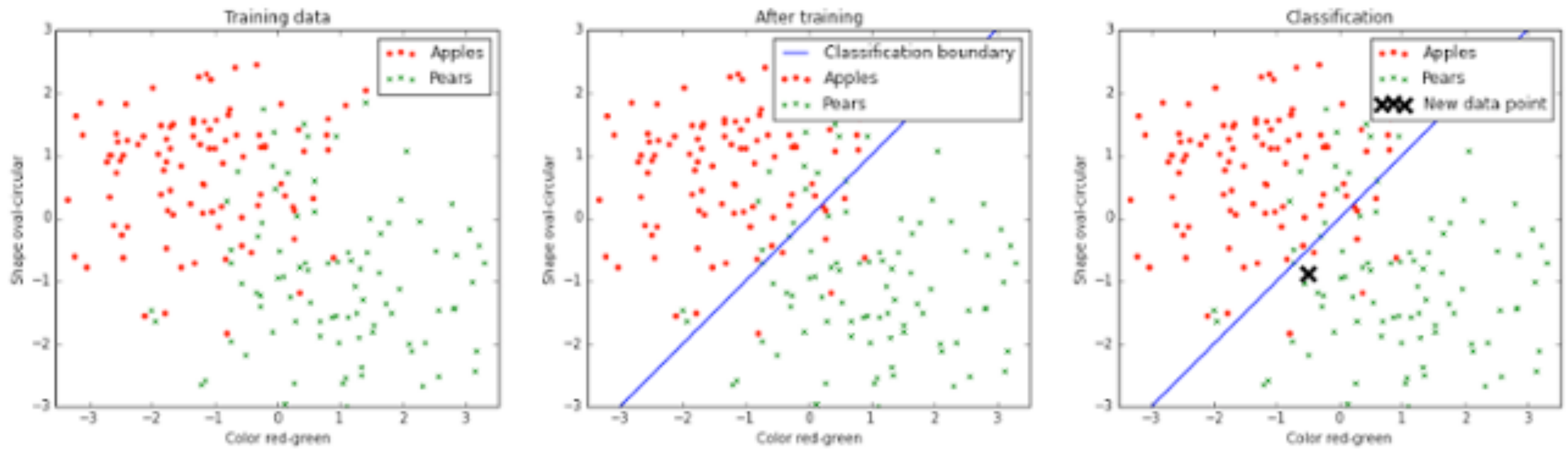
Logistic Regression (Binary Classification)

- Instead of predicting a real (continuous) value, we might want to predict a discrete value to classify certain data into categories such as cat/not_cat, spam/not_spam, pass/fail, S/M/L, cancerous/not_cancerous



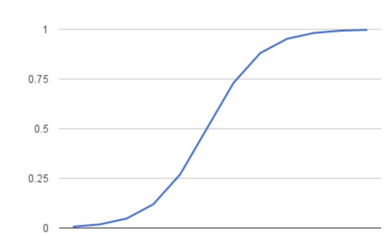
- Logistic regression is the relationship between features (x's) and the probability of a particular outcome (related to y's)
- We will use logistic regression to predict the probability of an event (e.g. whether it is a cat) for binary classification problems
- Logistic regression uses Maximum Likelihood Estimation – the set of coefficients (weights) for which the probability of getting the data we have observed is maximum

Linear Regression vs Logistic Regression



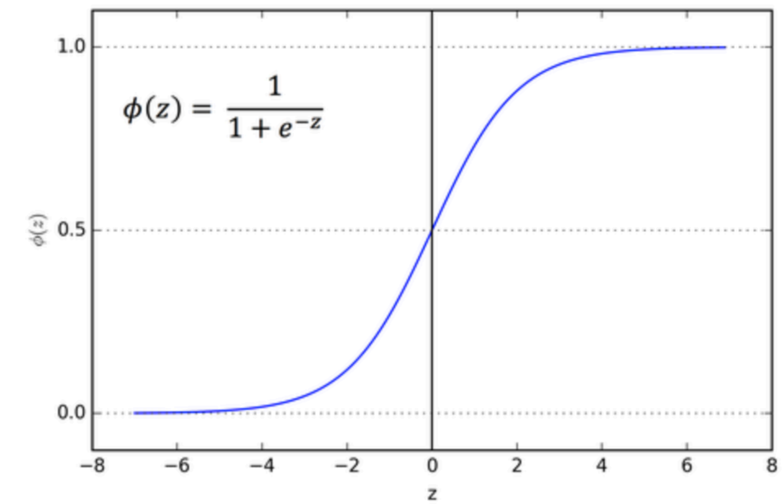
- Linear regression is used to predict a continuous value
  - Best fit line is a straight line, `y = wx + b`
- Logistic regression (or binary classification) is used to predict a discrete value
- A key difference from linear regression is that in logistic regression, the output value being modelled is a binary value (0 or 1) rather than a numeric value
- To restrict the output to be between 0 and 1 use a sigmoid (logistic) function
- Actually the output is a probability between 0 and 1, so the higher the probability, the more likely the value of the output is 1 (if probability > 0.5, then output=1)
  - The result of `wx + b` is fed to a logistic function to get `y`

$$y = \frac{1}{1 + e^{-z}} \qquad \text{where } z = wx + b$$

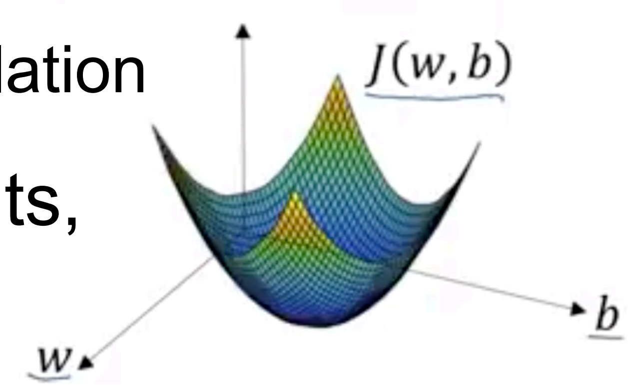


Sigmoid/logistic function

Logistic Regression cont.



- Analogous to linear regression
- Predicts the probability of an event/class
  - rounded to 0 or 1
- Finds weights, w for each feature, x
  - Positive implies positively correlated with outcome (e.g. `haveScale` and `beingReptile`)
  - Negative implies negatively correlated with outcome (e.g. `weeklyExerciseHours` and `havingHeartAttack`)
  - Absolute magnitude related to strength of the correlation
- Uses optimisation process to compute the weights, e.g. gradient descent to minimise the Log Loss



More on x's and y's

- In linear regression/classification, we may have a couple of scenarios:
  - One independent variable x affecting the final outcome (e.g. age affecting height)
  - More than one independent variable affecting the final outcome (e.g. age and sleep time affecting height)
  - More than one type of independent variable (continuous vs discrete) affecting the outcome (e.g. age, sleep time and gender affecting height)

Classification Method II

K Nearest Neighbours (kNN)

Animal Classification Example

Features | Label

Name	Egg-laying	Scales	Poisonous	Cold-blooded	Number legs	Reptile
Cobra	1	1	1	1	0	1
Rattlesnake	1	1	1	1	0	1
Boa constrictor	0	1	0	1	0	1
Chicken	1	1	0	1	2	0
Guppy	0	1	0	0	0	0
Dart frog	1	0	1	0	4	0
Zebra	0	0	0	0	4	0
Python	1	1	0	1	0	1
Alligator	1	1	0	1	4	1

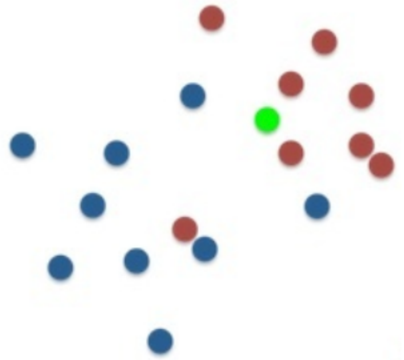
Distance matrix

	cobra	rattlesnake	boa constrictor	chicken	guppy	dart frog	zebra	python	alligator
cobra	--	0.0	1.414	2.236	1.732	1.732	2.236	1.0	1.414
rattlesnake	0.0	--	1.414	2.236	1.732	1.732	2.236	1.0	1.414
boa constrictor	1.414	1.414	--	2.236	1.0	2.236	1.732	1.0	1.414
chicken	2.236	2.236	2.236	--	2.449	2.0	2.0	2.0	1.0
guppy	1.732	1.732	1.0	2.449	--	2.0	1.414	1.414	1.732
dart frog	1.732	1.732	2.236	2.0	2.0	--	1.414	2.0	1.732
zebra	2.236	2.236	1.732	2.0	1.414	1.414	--	2.0	1.732
python	1.0	1.0	1.0	2.0	1.414	2.0	2.0	--	1.0
alligator	1.414	1.414	1.414	1.0	1.732	1.732	1.732	1.0	--

- Distance between different observations are recorded in a distance matrix
- Distance can be calculated using Euclidean or Manhattan, for example

1 Nearest Neighbour

- Using simple distance matrix for classification
- Simplest approach is probably one nearest neighbour
- Remember training data
- When predicting the label of a new example
  - Find nearest example in the training data
  - Predict the label associated with that example



Nearest Neighbour

	cobra	rattlesnake	boa constrictor	chicken	guppy	dart frog	zebra	python	alligator	Label
cobra	--	0.0	1.414	2.236	1.732	1.732	2.236	1.0	1.414	R
rattlesnake	0.0	--	1.414	2.236	1.732	1.732	2.236	1.0	1.414	R
boa constrictor	1.414	1.414	--	2.236	1.0	2.236	1.732	1.0	1.414	R
chicken	2.236	2.236	2.236	--	2.449	2.0	2.0	2.0	1.0	~R
guppy	1.732	1.732	1.0	2.449	--	2.0	1.414	1.414	1.732	~R
dart frog	1.732	1.732	2.236	2.0	2.0	--	1.414	2.0	1.732	~R
zebra	2.236	2.236	1.732	2.0	1.414	1.414				
python	1.0	1.0	1.0	2.0	1.414	2.0				
alligator	1.414	1.414	1.414	1.0	1.732	1.732				

Prediction: ~R, Actual: ~R ✓

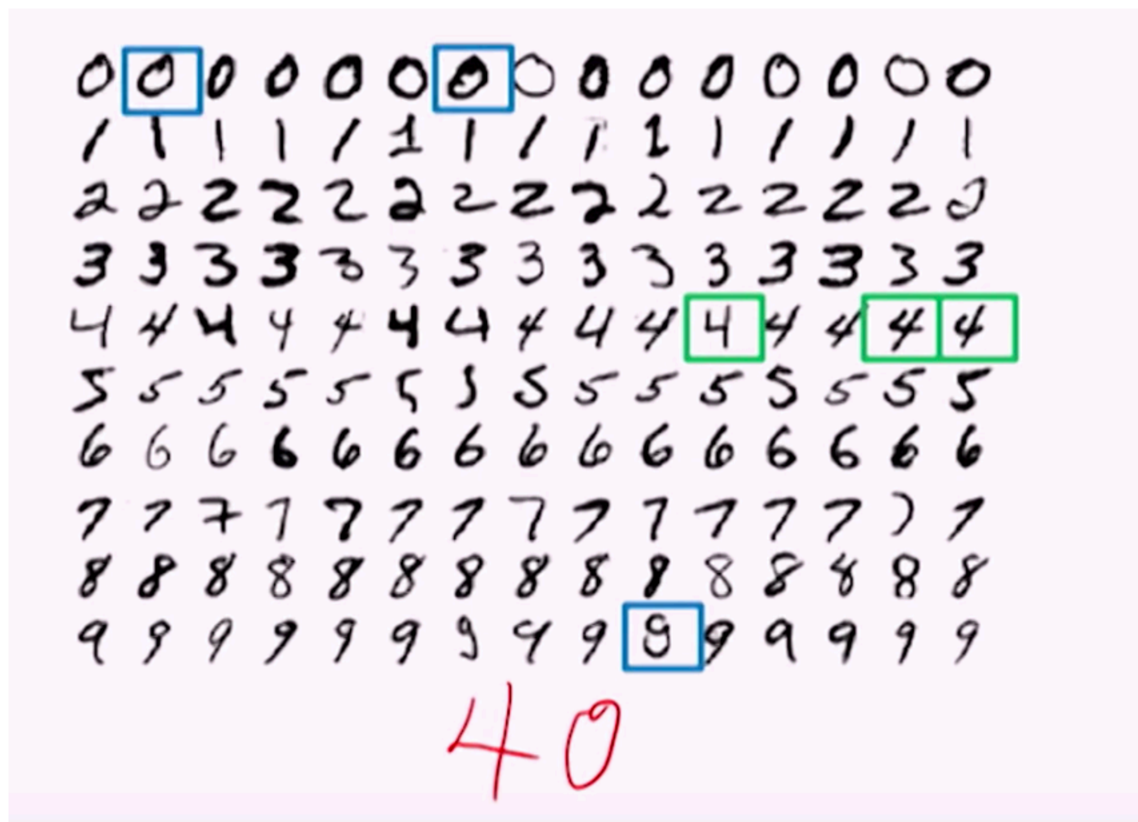
Prediction: R, Actual: R ✓

Prediction: ~R, Actual: R ✗

Choosing just one nearest neighbour could be problematic

K Nearest Neighbours (kNN)

- Take k nearest neighbours, k usually odd (if 2 classes), and take a vote on majority class



- Problem: if k is too high it becomes inefficient (memory intensive)

kNN Exercise

Lisa has lost gender information of one of her customers, and does not know whether to make a skirt or trousers. She is planning to throw a coin.

Can you help her to make a better decision using a KNN-classifier?

The customer who is missing gender information:
Gender ------, Waist **28**, Hip **34**

Let us use K = 3 nearest neighbors.

Fill in the table to calculate KNN.

Gender	waist (cm)	hip (cm)	distance	ranking number	belongs to the neighborhood (Yes or No)	gender of neighborhood member (empty if x_n not in the neighborhood)
Male	**28**	**32**	$(28-28)^2+(34-32)^2=4$			
Male	**33**	**35**	$(28-33)^2+(34-35)^2=26$			
Female	27	33				
Female	31	36				

Logistic Regression vs. kNN

- Feature weights in Logistic Regression give insight about variables – positively or negatively correlated to the outcome and strength
- kNN is comparatively slower than logistic regression when it comes to predicting
- kNN is has no explicit training (memorise all data and labels)
- kNN is a distance based technique while logistic regression is probability based

Summary

- Supervised learning (Classification)
    - logistic regression (binary classification)
    - k nearest neighbours (KNN)

Links

- Russell and Norvig Chapter 18
- A. Ng. Linear and Logistic Regression Cost Function (Videos) [LinearR1] [LinearR2] [LogisticR]
- J. Brownlee. Logistic Regression for Machine Learning [Link]
- A. Ng Deeplearning.ai Logistic Regression Working [Link] Loss & Cost function [Link], Training [Link]
- StatQuest with Josh Starmer (Videos) [Coefficients] [Max Likelihood] [Significance] [R-squared]