IAML 2017-2018

1. Linear and Logistic Regression

a. Lin. reg. predictive model has the form
$$y = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_d x_d$$
Give the form of the logistic model.

$$y = w^T x$$
logistic

weight vector

$$p(y=1|\underline{x}) = f(\underline{w}^T \underline{x})$$
input vector

probability that x belongs to class 1

the squashing function used e.g. $\sigma(z) = \frac{1}{1+\exp(-z)}$

For multi class, squashing fn = softmax

$$p(y=k|x) = \frac{\exp(w_k^T x)}{\sum_{i=1}^{K} \exp(w_i^T x)}$$

b. Weather data. Learn a model to predict next day's humidity based on the previous two weeks.
Would a linear model be sensible? What about logistic? Why?

predict real valued data

We need to predict a value so we use linear regression. Logistic Regression is used to classify the data into classes, i.e. humid or not humid.

c. Predict whether it rains or not.

Use logistic regression because it is a classification algorithm that splits the data into classes.

d. $y = w^T x$ predict the price of houses.

Atts → x_1 = avg. price of homes in the area
x_2 = no. of bedrooms in home.
x_3 = no. of years since house built.

$$\underline{w} = (w_0, w_1, w_2, w_3)^T = (40000, 0.5, 20000, -5000)^T$$

can you ignore the corresponding attribute?

Explain the validity in lin. regr of using weight values to determine significance of attributes.

e. softmax function for logistic regression

$$p(y=k \mid x) = \frac{\exp(w_k^T x)}{\sum_{i=1}^{K} \exp(w_i^T x)}$$

For two classes, $K=2$

$$p(y=1 \mid x) = \frac{1}{1+\exp(-z)} = \frac{\cancel{\exp(w_1^T x)}}{\cancel{\exp(w_1^T x)} + \exp(w_2^T x)}$$

$$= \frac{1}{1 + \exp(\underbrace{w_2^T x}_{\to -z \propto \sigma(z)})}$$

we have 10 features and classify to 2 outcomes.

overweight, $\neg$overweight, cancer, $\neg$cancer

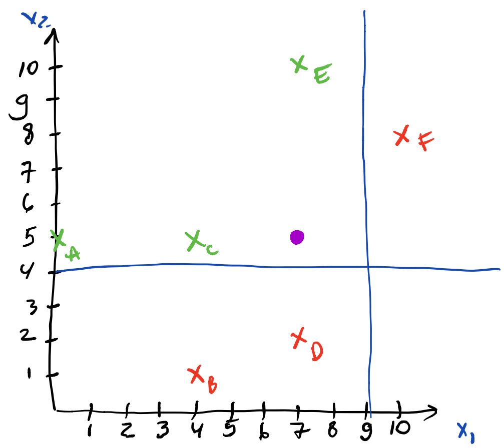
4-class classification problem $\to$ softmax function.

2. Decision Boundaries and Optimisation

a. Decision Boundaries

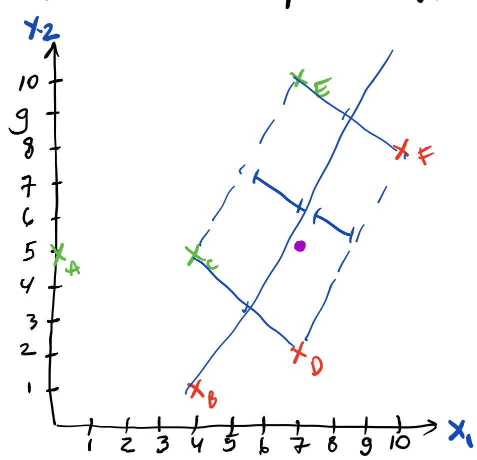
positive: A (0,5) C(4,5) E(7,10)
negative: B (4,1) D(7,2) F(10,7)

i. DB by DT



Decision tree

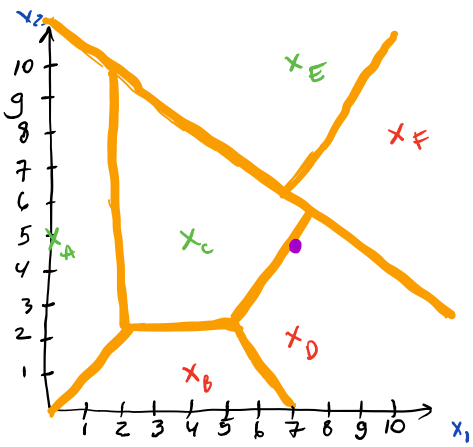
ii. Decision boundary with SVM



parallel to C, E and F, D

SV - B.

ii. DB with KNN



b. Optimisation

i. Gradient Descent

init w
while $E(w)$ unacceptably high
$g \leftarrow \partial E / \partial w$
$w \leftarrow w - \eta g$
end while
return w

ii. Learning rate

iii. Log regr model minimas

iii. Regularisation

3. PCA

a. Describe how to do dimensionality reduction

1. Compute the covariance matrix of $\underline{X}$
2. Get the eigen vectors /values
3. Sort eigenvalues s.t. $\lambda_1 \geq \lambda_2 \ldots$
4. Obtain the eigenvalues that rep. 95% of the var.
   ↳ corresponding eigenvectors
5. Project the data $\underline{X}$ into the principal components.

   project a to b → $b^T a$

b. Use KNN

Find nearest neighbor to x
Obtain their score/label values
Average over the values to get x's label

c. $f(\underline{z}) = \underline{w}^T \underline{z} + w_0$
Give suitable objective function to determine w.

$$\hat{w} = (\phi^T \phi)^{-1} \phi^T y$$

ϕ = training data matrix

Good for linear regression as dimensions are reduced.

d. Discuss the relative merits of the KNN and linear regression methods for this task.