1. Short answers
2. The advantages and drawbacks of leave-one out cross validation (LOOCV)?
3. What is “Bias-Variance Decomposition”? you can use formula to illustrate. What is the feature of over-fitting with regards to bias and variance?
4. Write down at least 2 machine learning models you learned, which are solved by MLE?
5. Describe at least 2 regularization methods for linear regression. What are theirs differences?
6. Decision Tree and Random Forest
7. . Describe the procedure of decision tree algorithm.
8. . Please write down the most commonly used measure information Gain.
9. . How begging and Random Forest introduce diversity in individual learners respectively? From bias-variance decomposition perspective, how they improve generalization performance?
10. Hierarchical Clustering

Apply the Hierarchical Clustering to the following distance matrix with single linkage. Write down each step of your clustering procedure.

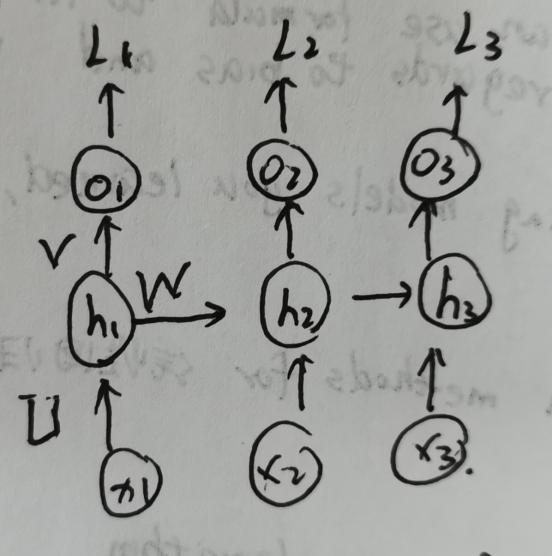
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F |
| A | — |  |  |  |  |  |
| B | 7 | — |  |  |  |  |
| C | 3 | 15 | — |  |  |  |
| D | 11 | 5 | 9 | — |  |  |
| E | 8 | 6 | 14 | 7 | — |  |
| F | 1 | 8 | 10 | 9 | 8 | — |

1. Naive Bayes Classifier

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 甲 | 乙 | 丙 | 丁 | y |
| 0 | 1 | 6 | 5 | +1 |
| 3 | 2 | 8 | 6 | +1 |
| 1 | 5 | 0 | 2 | -1 |
| 0 | 0 | 4 | 4 | +1 |
| 5 | 6 | 3 | 0 | -1 |
| 4 | 9 | 0 | 4 | -1 |
| 0 | 2 | 4 | 1 | +1 |
| 6 | 1 | 0 | 0 | -1 |

Suppose you are given 8 text “article” samples that are made up of only 4 unique characters “甲”、”乙”、”丙”、”丁”. The number of each character within the article is counted and listed in the above table.

1. Calculate the empirical conditional probability of each character for appearing in texts from each label. E.g.:P(word=甲|y=+1),P(word=甲|y=-1) . smoothing.
2. Predict the label of “甲甲乙丙丙丁丁丁”.
3. Back Propagation



Initialized =0, =  in each time step.

=  ←output

Loss  (each time). Total loss 

1. What is the total number of learnable parameters in this nn(there are no bias terms)?
2. Compute 
3. Compute , 

Tips: (tanhx)’ = 1-(tanhx) , 

1. SVM
2. min 

s.t. , i=1,2,...N

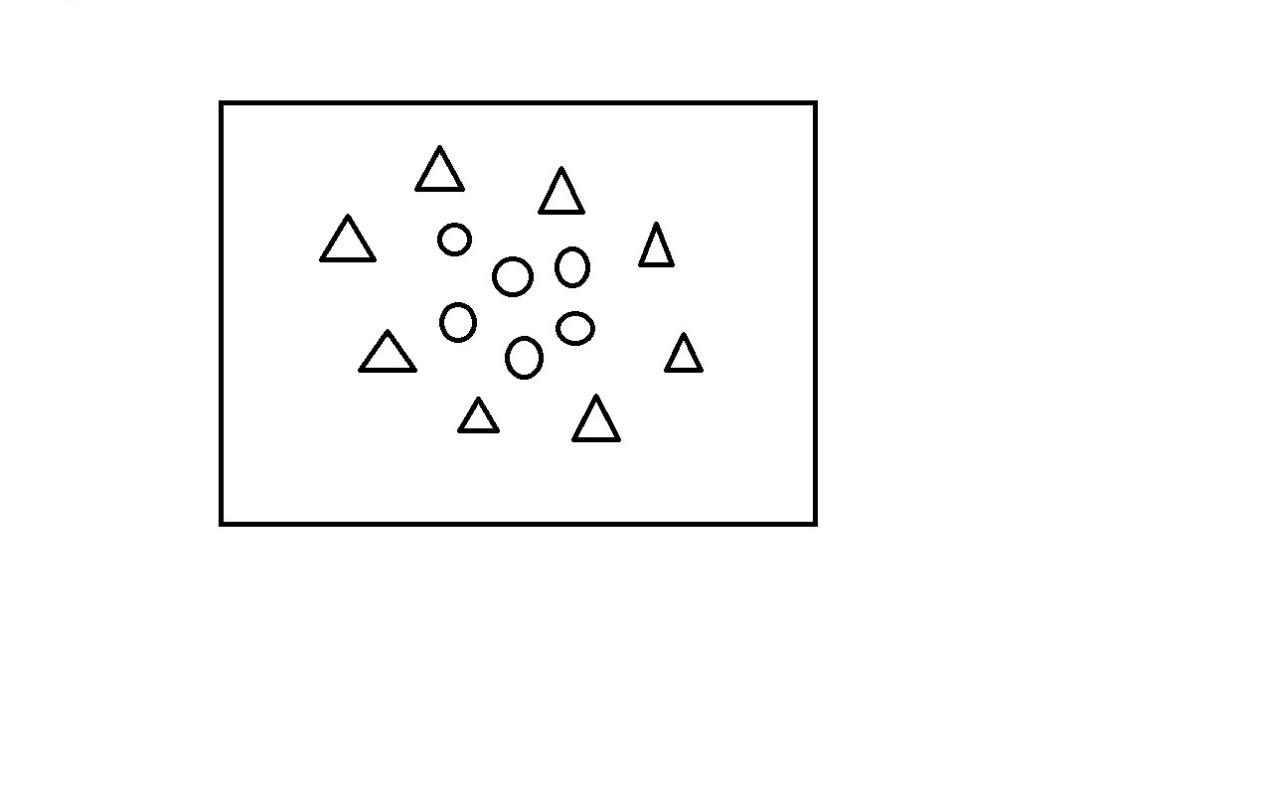
a) Derive its dual problem using the method of Lagrange multipliers and further simplify. The dual problem when at its saddle point to prove:



s.t.   is equivalent to the primal problem.

b) Suppose we have know the optimal value ,i=1,2,....N,derive w\* and b\*.

1. Given the centralized sample points as below. What kernel would you choose if train a hard margin SVM? Write down your K(xi,xj). Using your kernel,would the 决策边界 look like when the hard margin kernel SVM cenverges? Draw a possible 边界。



1. Logistic Regression

二分类 , xi=(xi1,...xid), y∈{0,1}

1. What does linear model wTx+b fit?
2. Derive posterior probability P(y=1|x) and P(y=0|x).
3. Derive the loss function (minimization objective) of LR model.
4. Gaussian Bayesian Classifiers

Given data set ,y∈Y={1,2,...K}

1. Please write down the Bayes optional classifier the min.... the misclassification error rate.
2. Suppose the samples in the K-th class are i.i.d sampled from normal distribution