142

1. iid assumption: independent and identically distributed(iid): A collection of random variables is iid if they all have the same probability distribution, and all are mutually independent.

2. supervised learning： classification:labels are nominal

- Binary Classification:{spam,not-spam}
- Multi-class classification:{0,1,2,3},{positive,negative,neutral}
- regression:labes are numeric(e.g. price of house)
- ranking problems(order a set of objects)
- reinforcement learning:
- output:sequence of actions. Individual actions arenot correct/incorrect but the overall policy is important
- no supervised output, but delayed reward
- On-line learning:train on one instance at a time: perceptron , contrasted with batch learning.
- Active learning- request labels for particular instances.

3. Unspervied learning- no labels provided at all at train time

- clustering
- image compression, bioinformatics.

4. Semi-supervised learning-used labeled as well as unlabeled data
5. Choosing a hypothesis too simple, fewer parameters to learn, but less powerful.

- Model has less variance:fewer changes with changing training data, model has more bias makes more assumptions.

6. Choosing a hypothesis too complex, more parameters to learn but more powerful, more variance and less bias. Called bias-variance tradeoff.

7. Bias represents estimation error(limitation of the model) Variance represent approximation error(limitations of the model family)

8. error(g)=bias+variacne.

9. over-fitting results in models that are more complex than necessary:after learning knowledge they "tend to learn noise". More complex models tend to have more complicated decision boundaries and tend to be more sensitive to noise, and missing examples.

10. Underfitting does not represents data well enough.

11. various evalutation measures exist in literature that can evaluate predictive performance. Most popular for classification:accuracy and error rate, precision recall and F-measure.

accuracy = $\frac { correct predictions }{number of test instances}$

error = 1 - accuracy

confustion matrix https://ucsc-courses.github.io/CMPS142-Spring2018/slides/Lecture1_2.pdf P35

12. k-fold cross validation

13. Discriminative learning:model the problem of text correction as a problem of learning from examples. Goal:learn directly how to make predictions. Model the problem of text correction as a problem of learning from examples. Goal: learn directly how to make predictions $P(Y|X)$ It focus on learning about not just the labels but also how instances were generated given those labels.Good at distinguishing between classes:learning boundaries.

14. Model the problem of text correction as that of generating correct sentences. Goal : learn a model of the data, use it to predict.$P(X,Y)=P(X,Y)P(Y)$It focus on learning about the labels give instances.Good at learning the underlying distribution of the data.

15. not all probabilistic models are Generative/Bayesian

16. Learning probabilisOc concepts – You can learn a concept which is a funcOon g:X→[0,1] – g(x) may be interpreted as the probability that x takes a certain value. E.g. probability that the label is “spam”.

17. Bayesian Learning: – use probabilisOc criterion to choose the hypothesis – The hypothesis can be determinisOc: e.g. a Boolean function, a rule – The criterion used to select the hypothesis is probabilisOc – It’s this process that makes the difference

18. 1.All probabilities are between 0 and 1. Probability of all possible world is 1 , the probability of a disjunction is give by $P(A\vee B)=P(A)+P(B)-P(A\wedge B)$

19. Joint probability: – Consider multiple variables and see how to behave together – Matrix of combined probabiliOes of a set of variables

20. Conditional probability: $P(A|B)=\frac {P(A\wedge B)}{P(B)}$

$P(A\wedge B) = P(A|B)* P(B)$

21. Independence: When two event do not affect each others' probabilities, we call them independent.$P(A\wedge B)= P(A)*P(B)$

$P(A|B)=P(A)$

22.