

6364

HW3.

1. False. less data with same complexity will cause underfit
2. False. when the feature space is larger, overfitting is more likely.
3. False. KNN is slower than Logistic Regression
4. True. KNN has faster training time
5. True. if the feature number is larger than the training samples
6. False. a proper learning rate is required
7. True
8. True
9. True
10. False. Cross validation does not prevent overfitting

2. ①

+	-	-
++	-	-
+	-	-

because large margin value, the SVM boundary will work

②

+	-	-
+	+	-
+	-	-

those three data point will depend the SVM boundary, remaining those point will cause a decision boundary.

3. 1. Naive assumption the conditional independence which simplify the computation

$$2. P(\text{buy}) = 9/14 \quad P(\text{youth}|\text{buy}) = \frac{3}{9} \quad P(\text{low}|\text{buy}) = \frac{3}{9}$$

$$P(\text{buy}) = 5/14 \quad P(\text{youth}|\text{buy}) = \frac{3}{5} \quad P(\text{low}|\text{buy}) = \frac{1}{5}$$

$$P(\text{student}|\text{buy}) = \frac{6}{9} \quad P(\text{fair}|\text{buy}) = \frac{6}{9}$$

$$P(\text{student}|\text{buy}) = \frac{1}{5} \quad P(\text{fair}|\text{buy}) = \frac{2}{5}$$

$$P(X|\text{buy}) = \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} = \frac{324}{6561} = 0.049$$

$$P(X|\text{buy}) = \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} = \frac{6}{625} = 0.0096 \quad \text{It could used for } P(X|\text{buy})$$

