Q1.1 False

Q1.2 True

Q1.3 False

Q1.4 True

Q1.5 True

Q1.6

文本

描述已自动生成

Q1.7

图片包含 文本

描述已自动生成

Q1.8

The original idea of svm is to use a linear separating hyperplane to create a classifier. Given training vectors xi, i=1,…,n, and a vector y is define as {-1, 1}, The two classes could be fully separated by a dotted line wx+b=0. Also, we want to decide the line with largest margin wx+b=+1, wx + b = -1 is maximize. Wx + b = +1 and wx + b = -1 are two support vectors. Thus, we call is a support vector.

Q1.9 True

Q1.10 False

Q1.11

The dual problem of svm is as follows:

文本, 信件

描述已自动生成

Q1.12

Q1.13 False

Q1.14 True

Q2.1

The distance between X\_i and x\_j in knn is as follows

L\_{p}(x\_i, x\_j) = (\sum\_{n}^{l=1}|x\_{i}^{l}-x\_{j}^{l}|)| ^ {p} )^{\frac{1}{p}}

When p is equal to infinite, we have:

L\_{\infty}(x\_i,x\_j) = max\_{l}|x\_i-x\_{j}|

Q2.2 不会，自己编一下

Q3.1

P(Y=1|x) = \frac{exp(-w^{T}x)}{1 + exp(-w^{T}x)}

P(Y=0|x) = \frac{1}{1+exp(-w^{T}x)}

The loss function of log-likelihood is as follows:

L(w) = \sum^{N}\_{i=1}[y\_i log(g(z)) + (1-y\_i)log(1-g(z))]

Use gradient descent, we can get optimal w as follows:

W\_{j+1} = w\_{j} + \alpha x^{T}(y-g\_{w}(x))

Q3.2

文本

描述已自动生成

Q 3.3

The likelihood function in Bayesian is as follows:

P(y|x)

If we use Gaussian Distribution into Bayesian, we have:

P(y | x, \alpha) = \frac{1}{\sqrt{2 \pi \sigma}}exp(-\frac{(x-\mu)^2}{2 \sigma^{2}})

The we have loss function as follows:

L(w,s) = \frac{1}{2 \sigma^{2}} \sum^{n}\_{i=1}(y\_i – x^{T}w)^2

We make the ·derivative ·of L(w,s) = 0,

Then we can solve the optimization w is :

W^{\*} = (x^{T}X)^{-1}x^{T}y

Q4.1

Remain roughly the same

Because regularization is to solve overfitting problem. Overfitting make training error decrease and testing error increase, when add L2 regularization, it can make training error increase, and testing error decrease, to make classifer more generalization. When j=0, it is same like no regularization, so training error is roughly the same. J=1 or 2 it make model more generalization, and testing error decrease, training error increase. J = 1 make \theta small, and j = 1 make \theta more smaller, it can make model more generalization

Q 4.2

increase. 原因同4.1

Q4.3

increase. 原因同4.1