В задаче 3 перепутаны местами пункты а и d. В задаче 4 нет явно выписанной двойственной задачи. Остальное верно.

Теоретическое задание 2

Выполнил: Бородин А.А.

***Задача 1.***

Пусть алгоритм модели , тогда логарифм правдоподобия выборки:

Данный логарифм правдоподобия можно использовать как функционал для линейного классификатора, в виде

Покажем, что алгоритм может возвращать числа вероятности

При логарифмической функции потерь

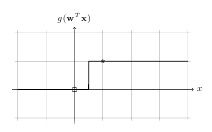
Запишем матожидание функции потерь в точке :

Полученная функция представляет собой логистические потери записанная для обозначения классов .

***Задача 2[[1]](#footnote-1).***

На первый взгляд, кажется, что проблема заключается в том, что решение оптимизационной задачи будет не единственно. Покажем более формально

Представим простой случай линейно разделимой выборки . В этом случае максимум правдоподобия будет достигаться ступенчатой функцией,



Любая другая функция, которая увеличивает вероятность не так резко, будет иметь меньшее значение правдоподобие. А любая другая ступенчатая функция будет так же приводить к максимуму правдоподобия. Это приводит к тому, что решение не единственно.

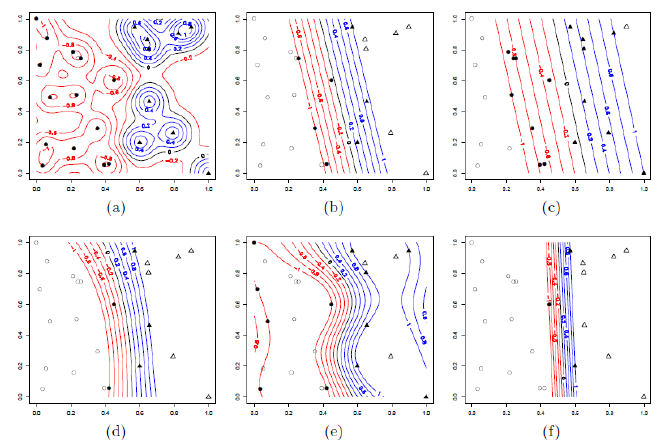
Чтобы избежать этой проблемы необходимо использовать регуляризацию. Т.е. от принципа максимума правдоподобия перейти к принципу максимума апостериорной вероятности, тем самым при оптимизации дополнительно использовать априорное распределение вектора параметров, который оцениваем.

Формулы пересчета значений параметров при оптимизации методом градиентного cпуска:

-для предложенной модели

-для обыной

***Задача 3.***



(1) с параметром C = 10 соответствует изображению (f)

(2) с параметром C = 1 соответствует изображению (b)

(3) с параметром C = 0.1 соответствует изображению (c)

(4) с параметрами γ = 1, C = 3 соответствует изображению (e)

(5) с параметрами γ = 10, C = 1 соответствует изображению (a) (d)

(6) с параметрами γ = 0.1, C = 15 соответствует изображению (d) (a)

***Задача 4***

Какую функцию потерь оптимизирует линейный метод опорных векторов?

Постройте задачу квадратичного программирования аналогичную SVM, но для оптимизации функции потерь

Выпишите двойственную к ней задачу.

Запишем функцию Лагранжа:

где - вектор переменных, двойственных к переменным .

Условия Куна-Таккера сводят задачу к поиску седловой точки функции Лагранжа:

***Задача 5[[2]](#footnote-2)***

Пусть удовлетворяет уравнению

Предположим

Следовательно,

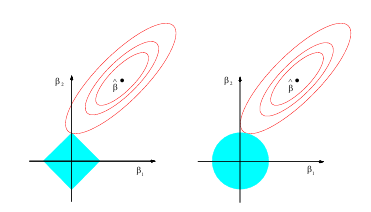
Таким образом является решением с наименьшей нормой.

***Задача 6[[3]](#footnote-3)***

Запишем функцию Лагранжа:

Известно, что если , то

Следовательно, минимизация приводит к решению исходной системы. Т.е. LASSO Тибширани эквивалентен l1-регуляризации для некоторого коэффициента регуляризации τ.



Оба метода l1 и l2 находят первую точку пересечения функции потерь и функции ограничения норм весов. В случае L1- регуляризации функции ограничения норм весов выражается (n-мерный ромб), в случае же L2- регуляризации функции ограничения норм весов имеет форму (n-мерной сферы). Если решением является угол ромба, тогда один из параметров обнуляется, в n-мерном пространстве возможны случаи когда не один параметр обнуляется. Сфера же не обладает подобным свойством.

1. https://www.cs.ubc.ca/~arnaud/cs340/HW5\_q2.pdf [↑](#footnote-ref-1)
2. http://www.seas.ucla.edu/~vandenbe/103/lectures/minnorm.pdf [↑](#footnote-ref-2)
3. [Elements of Statistical Learning [Trevor Hastie, Robert Tibshirani, Jerome Friedman]](http://statweb.stanford.edu/%7Etibs/ElemStatLearn/printings/ESLII_print10.pdf" \t "_blank" \o "Внешняя ссылка (откроется в новом окне)) [↑](#footnote-ref-3)