



## تکلیف سری سوم

مریم سعیدمهر  
شماره دانشجویی: ۹۶۲۹۳۷۳

### فهرست مطالب

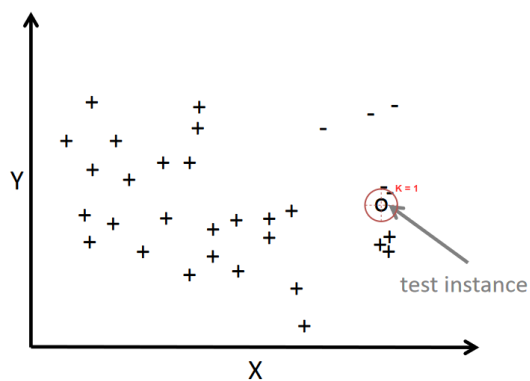
۲	۱ سوال اول
۲	۲ سوال دوم
۲	۱.۲ الف
۳	۲.۲ ب
۳	۳.۲ ج
۳	۳ سوال سوم
۵	۴ سوال چهارم
۵	۱.۴ الف
۶	۲.۴ ب
۶	۵ سوال پنجم
۸	۶ سوال ششم

۱ سوال اول

$$\begin{aligned}
 softmax(x)_i &= \frac{e^{x_i}}{\sum_j e^{x_j}} \\
 softmax(x+c)_i &= \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} \\
 &= \frac{e^{x_i} e^c}{\sum_j e^{x_j} e^c} \\
 &= \frac{e^{x_i} e^c}{e^c \sum_j e^{x_j}} \\
 &= \frac{e^{x_i}}{\sum_j e^{x_j}} \\
 &= softmax(x)_i \quad \blacksquare
 \end{aligned}$$

۲ سوال دوم

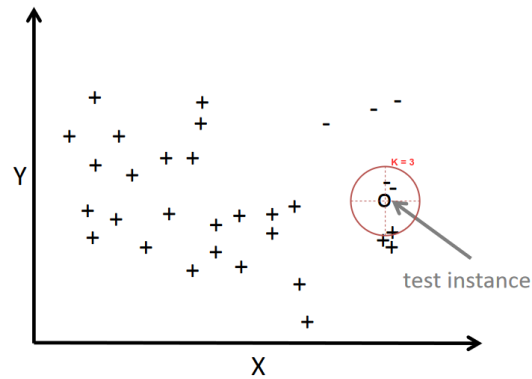
۱.۲ الف



شکل ۱:  $K = 1$

داده تست به کلاس (-) متعلق است.

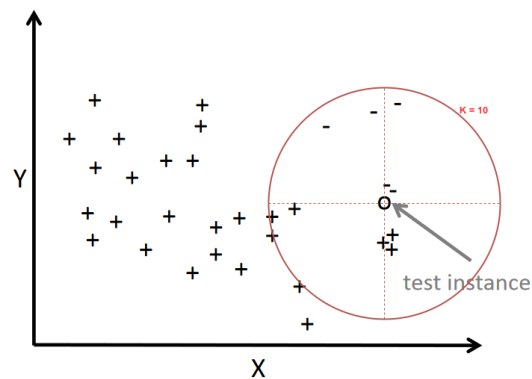
۲.۲ ب



شکل ۲:  $K = 3$

داده تست به کلاس (-) متعلق است.

۳.۲ ج



شکل ۳:  $K = 10$

داده تست به کلاس (+) متعلق است. در این حالت، در ۱۰ همسایگی داده تست ۵ داده متعلق به کلاس (+) داریم و ۵ داده متعلق به کلاس (-) پس صرفاً براساس فاصله نمیتوان قضاوت کرد لذا به هر داده وزن می‌دهیم و به این ترتیب داریم که: سه داده از کلاس (+) به داده تست نزدیک هستند و دو داده از این کلاس در فاصله دورتر و دو داده از کلاس (-) به داده تست نزدیک هستند و سه داده دیگر از این کلاس در فاصله دورتری از داده هستند، وزنه به سمت کلاس (+) سنگین تر است و لذا کلاس (+) را به داده تست منتسب میکنیم.

۳ سوال سوم

$X = (\text{color} = \text{green}, \text{legs} = 2, \text{height} = \text{tall}, \text{smelly} = \text{no})$

$$\text{Species} = f_{NB}(X) = \underset{\text{Species}}{\text{argmax}} P(\text{Species}|X) = \underset{\text{Species}}{\text{argmax}} P(\text{Species}) \prod_{j=1}^4 P(X_j|\text{Species})$$

• محاسبه احتمال prior :

$$P(\text{Species} = M) = \frac{\#M}{\#ID} = \frac{4}{8} = \frac{1}{2}$$

$$P(\text{Species} = H) = 1 - P(\text{Species} = M) = \frac{1}{2}$$

• محاسبه Likelihood (داده ها به شرط دانستن برجسب کلاس مستقل هستند<sup>۱</sup>):

$$P(\text{color} = \text{green} | \text{Species} = M) = \frac{2}{4} = \frac{1}{2}$$

$$P(\text{color} = \text{green} | \text{Species} = H) = \frac{1}{4}$$

$$P(\text{leg} = 2 | \text{Species} = M) = \frac{1}{4}$$

$$P(\text{leg} = 2 | \text{Species} = H) = \frac{4}{4} = 1$$

$$P(\text{height} = \text{tall} | \text{Species} = M) = \frac{1}{4}$$

$$P(\text{height} = \text{tall} | \text{Species} = H) = \frac{2}{4} = \frac{1}{2}$$

$$P(\text{smelly} = \text{no} | \text{Species} = M) = \frac{1}{4}$$

$$P(\text{smelly} = \text{no} | \text{Species} = H) = \frac{3}{4}$$

---

<sup>۱</sup> Naive Bayes Rule

• محاسبه برچسب نهایی داده X :

$$\begin{aligned}
 P(\text{Species} = M|X) &= P(\text{Species} = M) \prod_{j=1}^4 P(X_j|\text{Species} = M) \\
 &= \frac{1}{2} \times [P(\text{color} = \text{green}|\text{Species} = M) \times \\
 &\quad P(\text{leg} = 2|\text{Species} = M) \times \\
 &\quad P(\text{height} = \text{tall}|\text{Species} = M) \times \\
 &\quad P(\text{smelly} = \text{no}|\text{Species} = M)] \\
 &= \frac{1}{2} \times \left[ \frac{1}{2} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \right] \\
 &= \frac{1}{256}
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Species} = H|X) &= P(\text{Species} = H) \prod_{j=1}^4 P(X_j|\text{Species} = H) \\
 &= \frac{1}{2} \times [P(\text{color} = \text{green}|\text{Species} = H) \times \\
 &\quad P(\text{leg} = 2|\text{Species} = H) \times \\
 &\quad P(\text{height} = \text{tall}|\text{Species} = H) \times \\
 &\quad P(\text{smelly} = \text{no}|\text{Species} = H)] \\
 &= \frac{1}{2} \times \left[ \frac{1}{4} \times 1 \times \frac{1}{2} \times \frac{3}{4} \right] \\
 &= \frac{3}{64} = \frac{12}{256}
 \end{aligned}$$

$$\begin{aligned}
 \hat{\text{Species}} &= f_{NB}(X) = \arg_{\text{Species}} \max P(\text{Species}|X) \\
 &= \arg_{\text{Species}} \max \left\{ P(\text{Species} = M|X) = \frac{1}{256}, P(\text{Species} = H|X) = \frac{3}{256} \right\} \\
 &= H
 \end{aligned}$$

به این ترتیب  $X = (\text{color} = \text{green}, \text{legs} = 2, \text{height} = \text{tall}, \text{smelly} = \text{no})$  متعلق به کلاس H است

۴ سوال چهارم

۱.۴ الف

$$\begin{aligned}
 x^1 &= (4, 5), \quad y^1 = +1 \\
 x^2 &= (2, 3), \quad y^2 = -1 \\
 x^3 &= (1, 4), \quad y^3 = -1
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{L}(w, b, \alpha) &= \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^3 \alpha_i (y_i (x_i^T w + b) - 1) \\
 \frac{\partial \mathcal{L}}{\partial w} &= w - \sum_{i=1}^3 \alpha_i y_i x_i = 0 \implies w^* = \sum_{i=1}^3 \alpha_i y_i x_i = \alpha_1 \begin{bmatrix} 4 \\ 5 \end{bmatrix} - \alpha_2 \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \alpha_3 \begin{bmatrix} 1 \\ 4 \end{bmatrix} \\
 \frac{\partial \mathcal{L}}{\partial b} &= \sum_{i=1}^3 \alpha_i y_i = \alpha_1 - \alpha_2 - \alpha_3 = 0
 \end{aligned}$$

$$\alpha_1^*, \alpha_2^*, \alpha_3^* = \max_{\alpha_1, \alpha_2, \alpha_3} \mathcal{L}(w^*, b^*, \alpha); \quad \alpha_1, \alpha_2, \alpha_3 \geq 0, \sum_{i=1}^3 \alpha_i y_i = 0$$

$$\alpha_1^*, \alpha_2^*, \alpha_3^* = \max_{\alpha_1, \alpha_2, \alpha_3} \frac{-1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^3 \alpha_i$$

$$\begin{pmatrix} x_1^T x_1 = 41 & x_1^T x_2 = 23 & x_1^T x_3 = 24 \\ x_2^T x_2 = 13 & x_2^T x_3 = 14 & x_3^T x_3 = 17 \end{pmatrix}$$

$$\alpha_1^*, \alpha_2^*, \alpha_3^* = \max_{\alpha_1, \alpha_2, \alpha_3} \frac{-1}{2} [41\alpha_1^2 - 46\alpha_1\alpha_2 - 48\alpha_1\alpha_3 + 13\alpha_2^2 + 28\alpha_2\alpha_3 + 17\alpha_3^2] + \alpha_1 + \alpha_2 + \alpha_3$$

$$\alpha_1, \alpha_2, \alpha_3 \geq 0, \quad (\alpha_1 - \alpha_2 - \alpha_3 = 0 \Rightarrow \alpha_3 = \alpha_1 - \alpha_2)$$

$$\Rightarrow \alpha_1^*, \alpha_2^* = \max_{\alpha_1, \alpha_2} \frac{-1}{2} [41\alpha_1^2 - 46\alpha_1\alpha_2 - 48\alpha_1(\alpha_1 - \alpha_2) + 13\alpha_2^2 + 28\alpha_2(\alpha_1 - \alpha_2) + 17(\alpha_1 - \alpha_2)^2] + \alpha_1 + \alpha_2 + (\alpha_1 - \alpha_2)$$

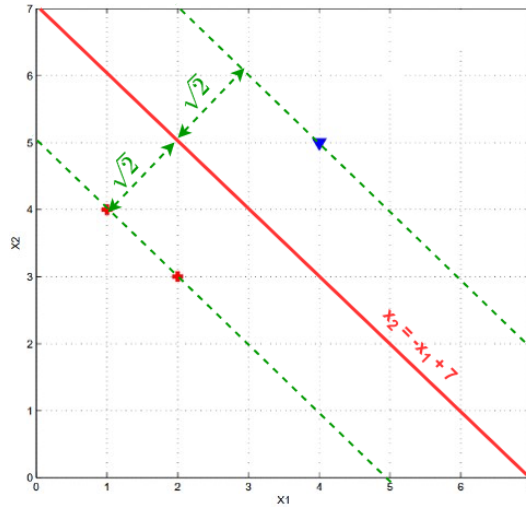
$$\left. \begin{aligned} \frac{\partial \mathcal{L}}{\partial \alpha_1} &= -10\alpha_1 + 2\alpha_2 + 2 = 0 \\ \frac{\partial \mathcal{L}}{\partial \alpha_2} &= -2\alpha_2 + 2\alpha_1 = 0 \Rightarrow \alpha_1 = \alpha_2 \\ \alpha_3 &= \alpha_1 - \alpha_2 \end{aligned} \right\} \Rightarrow \alpha_1 = \frac{1}{4}, \quad \alpha_2 = \frac{1}{4}, \quad \alpha_3 = 0$$

$$w^* = \sum_{i=1}^3 \alpha_i y_i x_i = \frac{1}{4} \begin{bmatrix} 4 \\ 5 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 0 \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad \blacksquare$$

$$m = \frac{2}{\|w\|} = 2\sqrt{2} \quad \blacksquare$$

$$b^* = y_2(x_2^T w + b^*) = 1 \Rightarrow b^* = \frac{-7}{2} \quad \blacksquare$$

ب ۲.۴

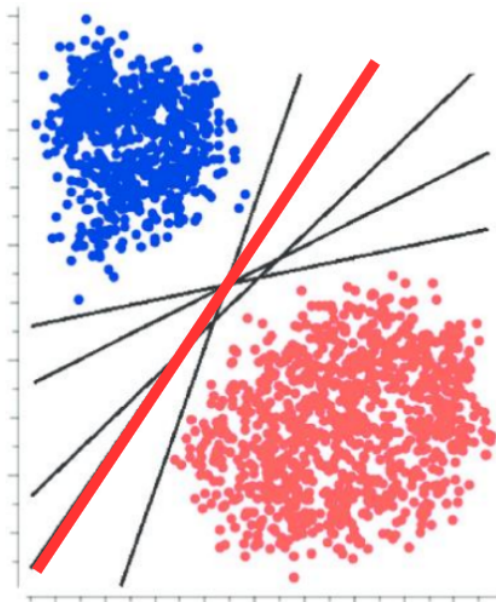


شکل ۴: رسم خط تصمیم. (هر سه داده بردارهای پشتیبان هستند.)

سوال پنجم ۵

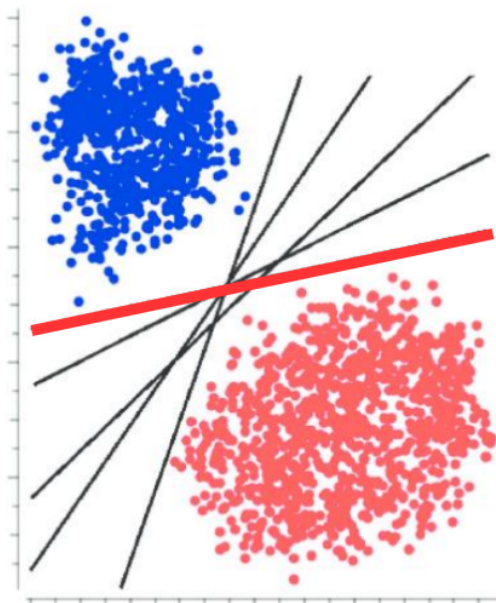
$$J(\theta) = \frac{1}{m} \sum_{i=1}^m [-y^i \log(h_\theta(x^i)) - (1 - y^i) \log(1 - h_\theta(x^i))]$$

۱. منتظم سازی  $\theta_0$  : خطای آموزشی تغییری نمیکند چون بایاس به صفر میل میکند. در واقع این ضریب میزان عرض از مبدأ را ست میکند و لذا منتظم سازی ضرایب نهایتاً باعث صفر شدن این ضریب میشود و به این معنی است که خط عرض از مبدأ صفر داشته باشد و از مبدأ مختصاً بگذرد پس از بین خطوط رسم شده در شکل خطی که به رنگ قرمز مشخص شده انتخاب میشود.



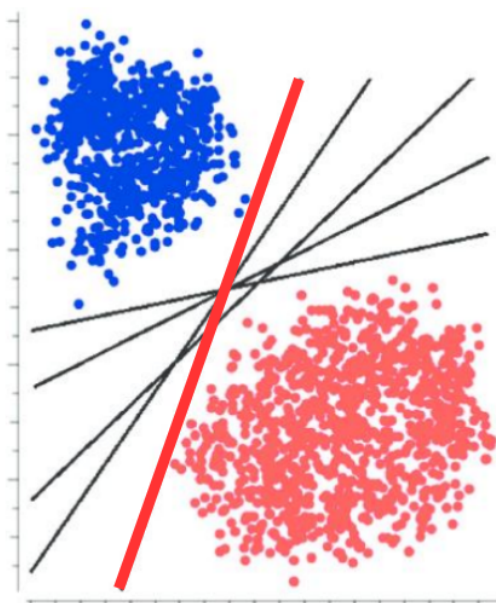
شکل ۵: منتظم سازی  $\theta_0$

۲. منتظم سازی  $\theta_1$  : در این حالت خطای آموزش افزایش مییابد زیرا وابستگی به  $x_1$  کاهش مییابد و مرز تصمیم به سمت افقی شدن میل میکند و این موجب رخدادن خطای دسته بندی است. در واقع در اثر منتظم سازی  $x_1$  مدل تمایل دارد خط قرمز رنگ زیر را از بین خطوط موجود انتخاب کند.



شکل ۶: منتظم سازی  $\theta_1$

۳. منتظم سازی  $\theta_2$  : در این حالت خطای آموزش افزایش مییابد زیرا وابستگی به  $x_2$  کاهش مییابد و مرز تصمیم به سمت عمودی شدن میل میکند و این موجب رخ دادن خطای دسته بندی است. در واقع در اثر منتظم سازی  $x_2$  مدل تمایل دارد خط قرمز رنگ زیر را از بین خطوط موجود انتخاب کند.



شکل ۷: منتظم سازی  $\theta_2$

## ۶ سوال ششم

با توجه به اینکه مدل روی داده های آموزش به مرور عملکرد خیلی خوبی داشته و میزان خطای آموزش آن با افزایش ابیایک ها کاهش یافته ولی روی validation set عملکرد خوبی نداشته میتوان فهمید که مدل overfit یا اصطلاحاً High variance است..

(آ) اضافه کردن ویژگی های جدید : تاثیری در بهبود مشکل واریانس مدل ما ندارد و فقط باعث کاهش بایاس میشود.

(ب) بزرگتر کردن مجموعه آموزشی : منجر به بهبود مدل خواهد شد و اگر تعداد داده ها به اندازه کافی زیاد شود میتواند مدل را از overfit شدن خارج کند و آن را به مدل بهینه و مطلوب تبدیل کند. (افزایش حجم داده ها باعث کاهش واریانس مدل میگردد)

(ج) بزرگتر کردن پارامتر منتظم سازی : منجر به ساده تر شدن مدل و کاهش واریانس میگردد و لذا باعث میشود مدل از overfit شدن خارج شود. (البته به شرط اینکه ضریب منتظم سازی بیش از حد بزرگ نشود که مدل را بیش از حد ساده یا اصطلاحاً underfit کند!)

نهایتاً میتوان گفت با افزایش حجم داده ها و یا به شکل مناسب tune کردن ضریب منتظم سازی میتوان مشکل این مدل (overfit بودن) را مرتفع ساخت.