

Machine Learning
Homework-2

پرسش 1:

* برای پاسخ به پرسش یک از فایل پایتون ضمیمه شده استفاده شود و این قالب پاسخ دهی را برای پرسش دو نیز در نظر بگیرید.

الف) مجموعه داده های جدول 1 را در نظر بگیرید: برای داده ها تابع جداساز برابر با $y = x$ است.

جدول 1.

ردیف	x	y	کلاس
1	2	1	C1
2	4	3	
3	3	2	
4	6	5	
5	5	4	
6	5	2	
7	6	3	
8	1	2	C2
9	3	4	
10	2	3	
11	5	6	
12	4	5	
13	2	5	
14	3	6	

این داده ها را برای دو کلاس با رنگ های متفاوت نمایش دهید.

حال تابع جدا ساز این داده ها را با استفاده از روش های زیر به دست آورید.

- بیز و نایو بیز
- رگرسیون خطی (با MSE و MAE) به عنوان تابع جریمه یا همان نرم یک و دو
- روش کلاس بندی رگرسیون لاجستیک
- روش SVM
- روش پرسپترون

ب) حال داده های جدول 2 را به داده های کلاس یک اضافه کنید.

جدول 2.

ردیف	x	y	کلاس
1	8	2	C1
2	10	4	

حال نتایج روش های بالا را با اضافه کردن این داده ها بررسی کنید. رفتار روش ها را با تغییر داده ها برای هر روش تحلیل کنید

روش SVM به این نوع داده های پرت (non-support vector datas) اصلا حساس نیست. روش logistic regression معمولاً حساسیت کمی نشان میدهد ولی در اینجا هیچ تغییری مشاهده نشد. در روش perceptron هم تغییری دیده نشد. در این سه روش کلاسه بندی ها به درستی انجام شده ولی روش naive bayes خطا دارد و علت آن این است که داده ها رو مستقل از هم در نظر میگیرد.

داده ها توسط روش MSE نسبت به روش MAE با دقت بیشتری پیشبینی شده ولی متأسفانه دقیق نمیدونستم این بخش رگرسیون رو چطور در کلاس بندی استفاده کنم و براش نمودار رسم کنم... ولی بطور کلی روش MSE چون توان دو داده ها رو در نظر میگیرد نسبت به روش نرم یک MAE حساسیت بیشتری به داده های پرت از خود نشان میدهد.

ج) علاوه بر داده های قسمت ب داده های جدول ۳ را نیز به کلاس یک اضافه کنید.

جدول 3.

ردیف	x	y	کلاس
1	2	7	C1
2	3	9	

دوباره روش های فوق را بر روی این روش ها انجام داده و توابع جداساز هر کدام را به دست آورید. تغییرات را تحلیل کنید.

این داده های پرت چون بطور کلی توزیع داده ها را به شکلی تغییر داده است که داده ها دیگر linear separable نیستند کلاسه بندی ها به شدت بد و با خطای بالایی عمل کردند. روش perceptron به نظر میرسد کلا عمل نکرده...
در اینجا شاید روش کرنل SVM (که متاسفانه فرصت نشد تا در کلاس به مبحث آن پرداخته شود) میتواندست خوب جواب دهد.

کد مرتبط با تمرین 1:

• Q1-HW2-Classification-Mahdavi.ipynb

پرسش 2:

جدول 4 را در نظر بگیرید:

جدول 14.

2D data		lable	Dual parameters
Xi1	Xi2	yi	α_i
4	2.9	1	0.414
4	4	1	0
1	2.5	-1	0
2.5	1	-1	0.018
4.9	4.5	1	0
1.9	1.9	-1	0
3.5	4	1	0.018
0.5	1.5	-1	0
2	2.1	-1	0.414
4.5	2.5	1	0

در اینجا ۱۰ تا داده همراه با پرچسب آنها داده شده است. فرض کنید آلفاهای داده شده، نتایج حاصل از حل مسئله دوگان روش SVM برای این داده ها باشد. در این صورت با استفاده از اطلاعات داده شده، ابرصفحه جدا ساز این داده ها را پیدا کنید. (منبع: زکی)

$$\text{hyperplan: } h(x) = w^T x_i - b$$

$$h(x) = (0.846, 0.3852) x - 3.54585$$

$$\text{bias: } b = y_i - w^T x_i = -3.54585$$

$$\text{weight: } w = \sum \alpha_i \cdot y_i \cdot x_i = (0.846, 0.3852)$$

فاصله نقطه X6 از این ابرصفحه چقدر است و آیا درون محدوده (Margin) کلاس بند می باشد؟

$$\varepsilon_i = 1 - y_i(W^T x_i + b) = -1.20112$$

$0 < \varepsilon_i < 1$ در محدوده margin قرار میگیرند پس ε_6 در خارج از آن قرار گرفته است.

همچنین با کمک ابرصفحه ی کلاس بند محاسبه شده، نقطه $z = (3, 3)^T$ را کلاس بندی کنید.

$$y = \begin{cases} +1 & \text{if } h(x) > 0 \\ -1 & \text{if } h(x) < 0 \end{cases}$$

چون 7.2394 بزرگتر از صفر است مربوط به کلاس یک است

کد مرتبط با تمرین 1:

Q2-HW2-Dual_SVM-Mahdavi.ipynb •

پرسش 3:

مسئله دوگان (Dual Solution) معادل با روش SVM زیر را پیدا کنید. (منبع: زکی 528)

$$\min_{w,b,E_i} \frac{\|w\|^2}{2} + C \sum_{i=1}^n E_i$$

$$y_i (W^T X_i + b) \geq 1 - E_i$$

$$E_i \geq 0$$

$$\min_{w,b,E_i} \left\{ \frac{\|w\|^2}{2} + C \sum_{i=1}^n (E_i)^k \right\}$$

$$y_i (w^T x_i + b) \geq 1 - E_i, \quad \forall x_i \in D$$

$$E_i \geq 0 \quad \forall x_i \in D$$

استفاده از Lagrangian:

$$\alpha_i (y_i (w^T x_i + b) - 1 + E_i) = 0 \quad \alpha_i \geq 0$$

$$\beta_i (E_i - 0) = 0 \quad \beta_i \geq 0$$

: Lagrangian

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n E_i - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1 + E_i) - \sum_{i=1}^n \beta_i E_i \quad A$$

با گرفتن مشتق جزئی نسبت به w ، b و E_i و همبستگی آن، آن را به Dual Lagrangian تبدیل می‌کنیم:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad \text{or} \quad w = \sum_{i=1}^n \alpha_i y_i x_i \quad \text{مشتق جزئی نسبت به } w :$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \quad \text{مشتق جزئی نسبت به } b :$$

$$\frac{\partial L}{\partial E_i} = C - \alpha_i - \beta_i = 0 \quad \text{or} \quad \beta_i = C - \alpha_i \quad \text{مشتق جزئی نسبت به } E_i :$$

جایگزینی مقادیر در معادله A:

$$L_{\text{dual}} = \frac{1}{2} w^T w - w^T \left(\underbrace{\sum_{i=1}^n \alpha_i y_i x_i}_w \right) - b \underbrace{\sum_{i=1}^n \alpha_i y_i}_0 + \sum_{i=1}^n \alpha_i + \underbrace{\sum_{i=1}^n (C - \alpha_i - \beta_i) E_i}_0$$

$$L_{\text{dual}} = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\max_{\alpha} L_{\text{dual}} = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$0 \leq \alpha_i \leq C, \quad \forall x_i \in D \quad \text{and} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

تعداد پارامترهای را مقایسه کنید و با فرض داشتن پاسخ بهینه ی مدل دوگان ، پاسخ بهینه مدل اولیه (Primal) را بیابید. (ابرفحه جداساز (Hyperplane) را بیابید).