تمرین 1 شبکه های عصبی – مازندرانیان - 830402066

سوال اول:Maximum Likelihood

فرض کنید $\{x_k\}, k = 1,2, \dots, N$ نمونه‌های مستقل از یکی از توزیع‌های زیر هستند. حداکثر درست‌نمایی θ را برای هر کدام به دست آورید:

a) $f(x_k;\theta) = \theta \exp(-\theta x_k)\ x_k \geq 0, \theta > 0$ Exponential Density

c) $f(x_k;\theta) = \sqrt{\theta} x_k^{\sqrt{\theta}-1}\ 0 \leq x_k \leq 1, \theta > 0$ Beta Density

برای Exponential Density تابع احتمال بصورت زیر است:

$$f(x\_k;\ \theta) = \theta * exp(-\theta * x\_k), x\_k >= 0, \theta > 0$$

تابع Likelihood مجموعه مشاهدات مستقل N ضرب احتمال تراکم آن ها است:

$$L(\theta) = \prod f(x\_k;\ \theta) = \prod \theta * exp(-\theta * x\_k) = \theta\text{^}N * exp(-\theta * \sum x\_k)$$

برای تخمین مقدار ماکسیمم، لگاریتم طبیعی تابع بالا را بدست میاوریم و سپس بر حسب تتا از آن مشتق می گیریم:

$$ln\ L(\theta) = N\ ln\ \theta - \theta * \sum x\_k$$

$$d/d\theta\ ln\ L(\theta) = N/\theta - \sum x\_k$$

مشتق را برابر صفر میگذاریم:

$$N/\theta - \sum x\_k = 0$$

$$\theta = N / \sum x\_k$$

بنابراین حداکثر درست نمایی برابر با میانگین است.

برای Beta Density:

$$f(x\_k;\ \theta) = \sqrt{\theta} * x\_k\text{^}(\sqrt{\theta} - 1), 0 <= x\_k <= 1, \theta > 0$$

$$L(\theta) = \prod f(x\_k;\ \theta) = \prod \sqrt{\theta} * x\_k\text{^}(\sqrt{\theta} - 1) = \theta\text{^}(N/2) * \prod x\_k\text{^}(\sqrt{\theta} - 1)$$

$$ln\ L(\theta) = (N/2)\ ln\ \theta + (\sqrt{\theta} - 1) * \sum ln\ x\_k$$

$$d/d\theta\ ln\ L(\theta) = N/(2\theta) + (1/(2\sqrt{\theta})) * \sum ln\ x\_k$$

از آنجایی که حساب کردن مقدار تتا در حالتی که مقدار مشتق را برابر صفر در نظر بگیریم مقداری سخت است، ولی از آنجایی که مقدار Likelihood وقتی بیشترین مقدار است که ضرب x_k بیشترین مقدار ممکن باشد، این حالت زمانی اتفاق میافتد که مقادیر x_k برابر با یک باشد بنابراین مقدار تتا برابر 1 خواهد بود.

سوال دوم: Regression

الف. مرحله به مرحله رابطه رگرسیون خطی با استفاده از واریانس، میانگین، کواریانس و correlation طبق مراحل زیر اثبات کنید:

۱. مفاهیم اولیه واریانس، میانگین، کواریانس و correlation را تعریف کنید.

۲. در رگرسیون خطی ساده، هدف یافتن رابطه‌ای خطی بین متغیر وابسته Y و متغیر مستقل X است. مدل به صورت زیر تعریف می‌شود:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

که در آن:

- Yمتغیر وابسته (پاسخ) است.
- Xمتغیر مستقل (پیش‌بینی‌کننده) است.
- β_0عرض از مبداء و β_1 شیب است.
- ε جمله خطا است که تفاوت بین مقادیر مشاهده شده و پیش‌بینی شده Y را نشان می‌دهد.

با استفاده از رابطه بالا مجموع مربعات باقی‌مانده[1] را کمینه کنید.

۳. استخراج شیب (β_1) و عرض از مبدأ (β_0) بر حسب کوواریانس و واریانس

۴. چگونه می‌توانیم β_1 را بر حسب ضریب همبستگی (ρ) و انحراف معیارها بیان کنیم؟ این رابطه چه چیزی را درباره ارتباط بین شیب خط رگرسیون و قدرت همبستگی بین متغیرها نشان می‌دهد؟

۵. فرمول نهایی برای شیب (β_1) بر حسب کوواریانس و واریانس چیست؟ فرمول نهایی برای عرض از مبدأ (β_0) چگونه با میانگین‌های X و Y مرتبط است؟

۶. این روابط چه بینشی درباره ارتباط بین ضرایب رگرسیون و معیارهای آماری مانند واریانس، کوواریانس و همبستگی به ما می‌دهند؟ چگونه این روابط به ما در درک بهتر خط رگرسیون تخمین زده شده کمک می‌کنند؟

ب. جدول زیر مربوط به یک مسئله رگرسیون خطی ساده است یا استفاده از فرمول‌های بدست آمده در قسمت قبل واریانس، میانگین، کواریانس و correlation را گزارش کنید.(کد نویسی با پایتون آنجام شود.)

i	x_i	y_i
1	16	46
2	27	80
3	11	36
4	20	52
5	30	98
6	25	75
7	5	10
8	24	70
9	21	64
10	10	30

سوال 1:

واریانس: معیار گستردگی یک دیتاست

کوواریانس: معیار اندازه گیری شباهت دو ویژگی با هم

کورلیشن: معیار اندازه گیری استاندارد رابطه ی بین دو متغیر

سوال 2:

$$SSE = \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\partial SSE/\partial \beta_0 = -2\sum (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\partial SSE/\partial \beta_1 = -2\sum X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\sum Y_i = n\beta_0 + \beta_1 \sum X_i$$

$$\sum X_i Y_i = \beta_0 \sum X_i + \beta_1 \sum X_i^2$$

$$\beta^1 = \sum \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\square} \sum (X_i - \bar{X})^2 \, \beta_0 = Y - \beta_1 X$$

سوال 3:

$$\beta_1 = Cov(X,Y) \,/\, Var(X)$$

$$\beta_0 = Y - \beta_1 X$$

$$r = Cov(X,Y) \,/\, \sqrt{(Var(X) * Var(Y))}$$

$$\beta_1 = r * (SD(Y) \,/\, SD(X))$$

منظور از SD همان انحراف از معیار است.

سوال 4:

به طور کلی، رابطه بین این مفاهیم به صورت زیر بیان می‌شود:

$\beta_1 = \rho * (\sigma y / \sigma x)$

در این رابطه:

- **β_1:** شیب خط رگرسیون است که نشان می‌دهد با یک واحد افزایش در متغیر مستقل(X) ، به طور متوسط چقدر متغیر وابسته (Y) تغییر می‌کند.
- **ρ:** ضریب همبستگی پیرسون است که میزان و جهت رابطه خطی بین دو متغیر را نشان می‌دهد. مقدار آن بین -1 تا 1 متغیر است.
- **σy:** انحراف معیار متغیر وابسته (Y) است.
- **σx :**انحراف معیار متغیر مستقل (X) است.

این رابطه چندین نکته مهم را نشان می‌دهد:

- **جهت رابطه :**علامت β_1 (مثبت یا منفی) با علامت ρ یکسان است. یعنی اگر همبستگی مثبت باشد، شیب نیز مثبت و اگر همبستگی منفی باشد، شیب نیز منفی خواهد بود.

- **قدرت رابطه :**مقدار مطلق β_1 نشان می‌دهد که با یک واحد تغییر درX ، به طور متوسط چقدر تغییر در Y انتظار می‌رود. هرچه مقدار مطلق β_1 بزرگتر باشد، شیب خط تندتر و رابطه بین دو متغیر قوی‌تر است.
- **نقش انحراف معیارها :**انحراف معیارها نشان می‌دهند که داده‌ها چقدر پراکنده هستند. اگر انحراف معیار متغیر وابسته بزرگتر باشد، تغییرات در Y نسبت به تغییرات در X حساس‌تر خواهد بود و در نتیجه شیب خط تندتر خواهد بود. به عبارت دیگر، هرچه پراکندگی داده‌ها بیشتر باشد، تاثیر تغییر در متغیر مستقل بر متغیر وابسته بیشتر خواهد بود.

سوال 5:

عرض از مبدأ (β_0) نقطه‌ای است که خط رگرسیون محور Y را قطع می‌کند. رابطه بین عرض از مبدأ و میانگین‌های X و Y به صورت زیر است:

$$\beta_0 = \bar{y} - \beta_1 * \bar{x}$$

تفسیر فرمول‌ها

- **شیب (β_1):** نشان می‌دهد که با یک واحد افزایش در متغیر مستقلX ، به طور متوسط چقدر متغیر وابسته Y تغییر می‌کند. اگر کوواریانس مثبت باشد، شیب نیز مثبت و رابطه مستقیم بین دو متغیر وجود دارد. اگر کوواریانس منفی باشد، شیب نیز منفی و رابطه عکس بین دو متغیر وجود دارد. واریانس متغیر مستقل در مخرج کسر نشان می‌دهد که پراکندگی داده‌های X چقدر بر شیب تأثیر می‌گذارد.
- **عرض از مبدأ (β_0):** مقدار پیش‌بینی‌شده برای متغیر وابسته Y هنگامی است که متغیر مستقل X برابر با صفر باشد. این نقطه لزوماً روی داده‌های واقعی قرار ندارد و صرفاً یک نقطه مرجع برای خط رگرسیون است.

رابطه بین شیب خط رگرسیون(β_1) ، ضریب همبستگی(ρ) ، و انحراف معیارهای متغیرهای مستقل (X) و وابسته (Y) به صورت زیر است:

$$\beta_1 = \rho * (\sigma y / \sigma x)$$

با توجه به تعریف ضریب همبستگی پیرسون که بر اساس کوواریانس و انحراف معیارها تعریف می‌شود، می‌توانیم فرمول را به صورت زیر بازنویسی کنیم:

$$\beta_1 = cov(X,Y) / var(X)$$

در این رابطه:

- $\boldsymbol{cov(X,Y)}$**:** کوواریانس بین متغیرهای X و Y است. کوواریانس نشان‌دهنده میزان تغییرات همزمان دو متغیر نسبت به میانگین‌هایشان است.
- $\boldsymbol{var(X)}$**:** واریانس متغیر مستقل X است. واریانس نشان‌دهنده پراکندگی داده‌ها حول میانگین است.

سوال 6:

```
[1] import numpy as np
```

```
X = [16, 27, 11, 20, 30, 25, 5, 24, 21, 10]
Y = [46, 80, 36, 52, 98, 75, 10, 70, 64, 30]

mean_X = np.mean(X)
mean_Y = np.mean(Y)

var_X = np.var(X)
var_Y = np.var(Y)

cov_XY = np.cov(X, Y)[0, 1]

corr_XY = np.corrcoef(X, Y)[0, 1]

print("Mean of X:", mean_X)
print("Mean of Y:", mean_Y)
print("Variance of X:", var_X)
print("Variance of Y:", var_Y)
print("Covariance of X and Y:", cov_XY)
print("Correlation of X and Y:", corr_XY)
```

```
Mean of X: 18.9
Mean of Y: 56.1
Variance of X: 60.08999999999999
Variance of Y: 626.89
Covariance of X and Y: 213.1222222222222
Correlation of X and Y: 0.9882674062434096
```

سوال سوم: طبقه بندی

در این سوال، یک طبقه‌بند طراحی کنید که بتواند دو کلاس متفاوت (دو تیم فوتبال منچستریونایتد و چلسی) را با استفاده از دیتاست داده‌شده تشخیص دهد. برای طبقه‌بندی، می‌توانید میانگین رنگ در هر عکس را محاسبه کنید و سپس مقدار به‌دست‌آمده را با رنگ‌های آبی و قرمز مقایسه نمایید. این طبقه‌بند را روی دیتاست داده‌شده تست کنید. ماتریس Confusion را گزارش دهید و مقادیر accuracy ، precision، و recall را محاسبه کرده و نتایج هر کدام را توضیح دهید.

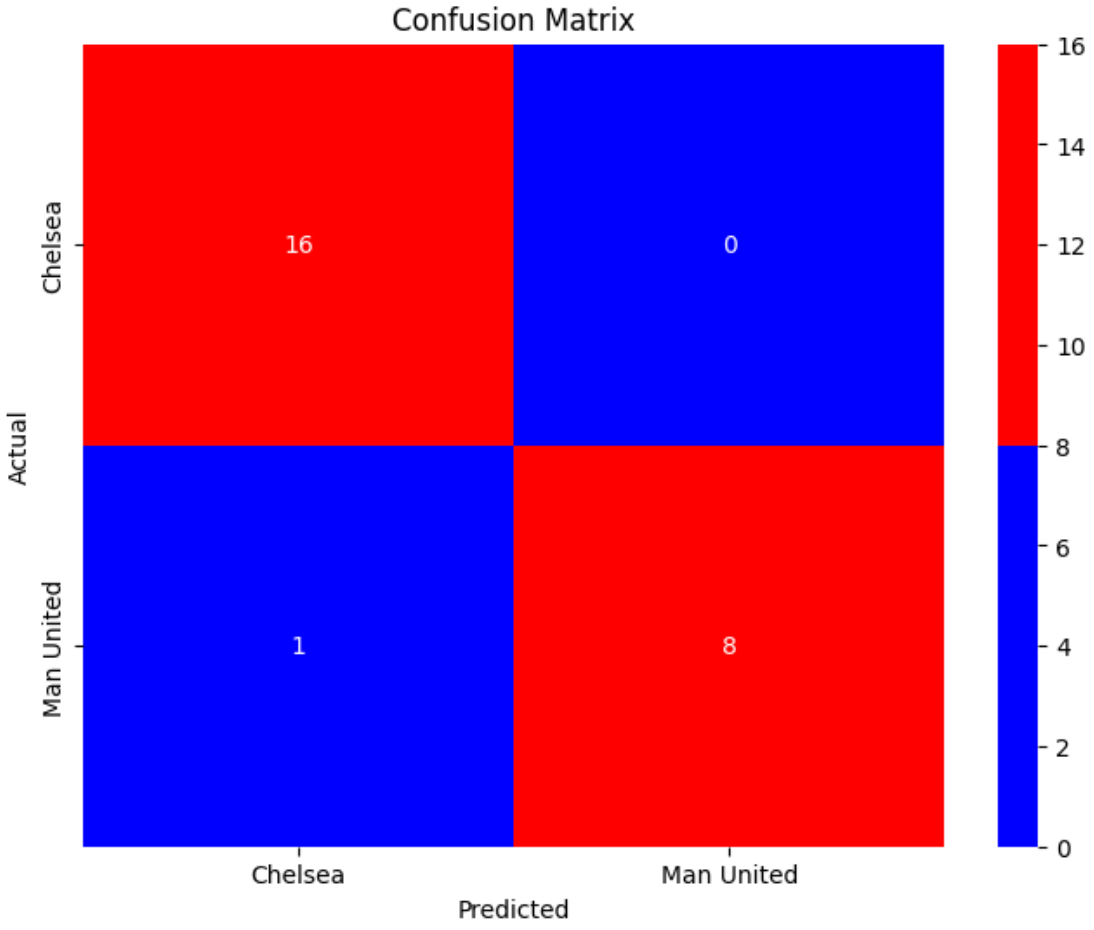
نکته: بدین منظور میتوان از میانگین کانال های عکس RGB استفاده کرد.

برای این طبقه بند از مدل Logistic Regresison استفاده شده است. ابتدا عکس های تیم های چلسی و منچستر یونایتد را جداگانه label های صفر و یک در نظر گرفتیم و میانگین کانال های RGB را به عنوان Feature در نظر گرفتیم. سپس داده ها را به 80 درصد آموزشی و 20 درصد تست تقسیم کردیم و مدل را Fit و سپس داده های تست را Predict کردیم.

```
print("F1-score:", f1)
```

```
Accuracy: 0.96
Precision: 1.0
Recall: 0.8888888888888888
F1-score: 0.9411764705882353
```

دقت مدل



ماتریس confusion

سوال چهارم: بررسی برازش[1] توابع مختلف

ابتدا با توجه به کد زیر داده‌های مربوطه را تولید کنید:

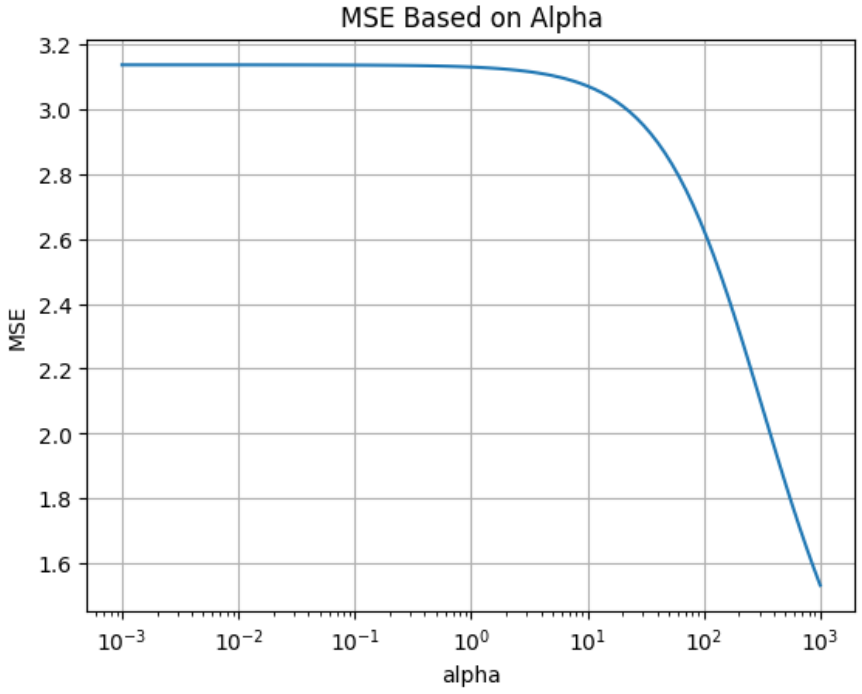
```
x = np.arange(-10, 10, 0.2)
y = 2 * cos(x) / -pi + 2 * sin(2 * x) / (2 * pi) + 2 *
cos(3 * x) / (-3 * pi)
```

سپس داده‌ها را با استفاده از K-fold به k بخش تقسیم کنید.

الف. در ابتدا سعی کنید توابع زیر را برازش کنید و مقادیر میانگین MSE برای هر یک از موارد را گزارش دهید.

1. Linear Regression
2. Polynomial Regression
3. Ridge Regression (L2 Regularization)

ب. در این بخش ضریب regularization رو در ridge Regression را بهینه و گزارش کنید. (در این بخش رسم نمودار MSE در فرایند بهینه سازی نمره اضافه به همراه دارد.)



در این، هدف برازش یک منحنی به داده‌های داده شده با استفاده از رگرسیون خطی، رگرسیون چند جمله‌ای و رگرسیون ریج بود. برای ارزیابی عملکرد هر مدل، از اعتبارسنجی متقاطع k-fold با k=5 استفاده شد و میانگین مربعات خطا (MSE) به عنوان معیار ارزیابی محاسبه شد.

رگرسیون خطی :یک مدل خطی ساده برای برازش داده‌ها استفاده شد.

رگرسیون چند جمله‌ای :با استفاده از ویژگی‌های چند جمله‌ای درجه 3، یک مدل غیرخطی به داده‌ها برازش داده شد.

رگرسیون ریج :برای جلوگیری از بیش‌برازش، از رگرسیون ریج با مقدار آلفا 0.5 استفاده شد.

در نهایت، با استفاده ازGridSearchCV ، بهترین مقدار آلفا برای رگرسیون ریج پیدا شد و نمودار MSE بر اساس آلفا رسم شد. نتایج نشان داد که رگرسیون چند جمله‌ای و رگرسیون ریج عملکرد بهتری نسبت به رگرسیون خطی در برازش منحنی به داده‌های داده شده دارند. در کد ارائه شده، مقدار آلفا 0.5 برای رگرسیون ریج به صورت **دلخواه** انتخاب شده است. در واقع، بهترین مقدار آلفا برای هر مسئله می‌تواند متفاوت باشد و به داده‌های مورد استفاده بستگی دارد. برای پیدا کردن بهترین مقدار آلفا، می‌توان از روش‌هایی مانند GridSearchCV استفاده کرد که در بخش آخر کد ارائه شده نیز از آن استفاده شده است.با اجرای GridSearchCV، بهترین مقدار آلفا برای این مسئله خاص پیدا شده و در متغیر best_alpha ذخیره شده است. بنابراین، اگرچه در بخش اول کد از آلفا 0.5 استفاده شده است، اما در نهایت بهترین مقدار آلفا با استفاده از GridSearchCV پیدا شده و می‌توان از آن برای برازش مدل نهایی استفاده کرد.