

تمرینات سری چهارم یادگیری ماشین
استاد بابا علی

ارشیا یوسفی نژاد
610302085

سوال اول

$$\begin{matrix} x_1 \rightarrow (2) \\ x_2 \rightarrow (-3) \end{matrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

$$\mu_1 = \frac{2+1+(-1)+(-1)+0+(-1)+(-1)+(-2)}{7} = -0.5$$

$$\mu_2 = \frac{-3+2+1+1+(-3)+(-1)+(-2)}{7} = -0.5$$

$$\text{Cov}(x_1, x_1) = \frac{1}{7} (2.5^2 + 1.5^2 + (-0.5)^2 + (-0.5)^2 + 0.5^2 + (-1.5)^2 + (-1.5)^2) = 2$$

$$\text{Cov}(x_1, x_2) = \frac{1}{7} ((-2.5)^2 + 3.5^2 + 1.5^2 + 1.5^2 + (-2.5)^2 + (-0.5)^2 + (-1.5)^2) = 4$$

$$\text{Cov}(x_2, x_1) = \text{Cov}(x_1, x_2) = \frac{3}{7}$$

$$\text{cov-matrix} = \begin{pmatrix} 2 & \frac{3}{7} \\ \frac{3}{7} & 4 \end{pmatrix}$$

$$\begin{vmatrix} 2-\lambda & \frac{3}{7} \\ \frac{3}{7} & 4-\lambda \end{vmatrix} = 0 \Rightarrow (2-\lambda)(4-\lambda) - \left(\frac{3}{7}\right)^2 = 0 \Rightarrow \begin{cases} \lambda_1 = 3 - \frac{\sqrt{58}}{7} \approx 1.911 \\ \lambda_2 = \frac{\sqrt{58} + 21}{7} \approx 4.087 \end{cases}$$

$$(C - \lambda_1 I) v_1 = 0 \Rightarrow v_1 = \begin{pmatrix} -0.201 \\ -0.998 \end{pmatrix}$$

$$(C - \lambda_2 I) v_2 = 0 \Rightarrow v_2 = \begin{pmatrix} -0.998 \\ 0.201 \end{pmatrix}$$

از آنجایی که λ_1 بزرگتر است پس

$$X v_1 = [2.926, -2.16, -0.778, -0.778, -0.979, 3.34, 1.18, 2.361]$$

$$X v_1 \cdot v_1^T = X v_1 \cdot (-0.201 \quad -0.9799) \Rightarrow$$

$$\begin{bmatrix} -0.51 & -2.484 \\ 0.434 & 2.11 \\ 0.156 & 0.7626 \\ 0.156 & 0.7626 \\ 0.1969 & 0.969 \\ -0.671 & -3.272 \\ -0.937 & -1.156 \\ -0.974 & -2.313 \end{bmatrix}$$

سوال دوم

برای حل این سوال از یک کلاس استفاده شده. که داخل کلاس توابع زیر تعریف شده است.

`mean_squared_error_reconstruction`

وظیفه‌ی این تابع این است روی دادگان ابتدا کاهش بعد را اعمال کرده سپس دوباره این کاهش بعد را بازسازی میکند و در پایان هم خطا را بصورت MSE روی دادگان بازسازی شده و دادگان اصلی حساب میکند.

`train_pca_reconstruction`

وظیفه این تابع این است که کاهش بعد را روی ابعاد مختلف امتحان میکند

`evaluate_pca_reconstruction`

این تابع بهترین ابعاد از pca رو گرفته (یعنی ابعادی که کمترین میزان حذف ویژگی داشته ایم) و روی دادگان تست MSE را حساب میکند

`show_image`

این تابع تصاویر بازسازی شده از pca و تصاویر اصلی را نشان میدهد. (نمودار تصاویر هم ضمیمه شده است)

`box_plot`

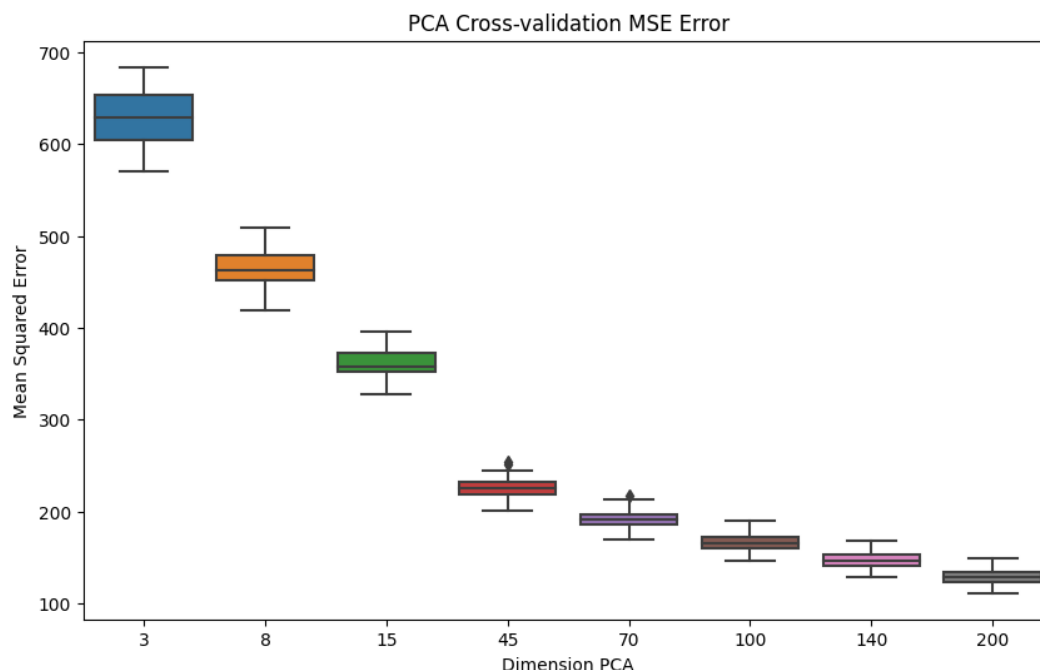
در اینجا نتایج cross validation ها را بصورت نمودار جمعی نمایش داده شده.

و در پایان از کلاس PCAModel یک object ای درست میکنیم. مقدار ابعاد گفته شده در تمرین را بهش داده و توابع مورد نیاز را فراخوانی میکنیم.

بخش اول

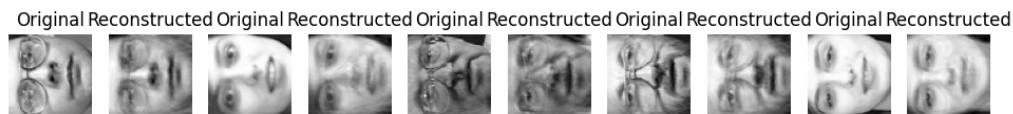
کاهش بعد به بعد ۱۰۰ خطا کاهش چشم گیری گرفته و از بعد صد به بعد سرعت کاهش خطا کم شده. (خطای زمانی که به بعد ۲۰۰ تغییر میدهیم با زمانی که به بعد ۱۰۰ تغییر میدهیم نا چیز است) (همانطور که در نمودار ضمیمه شده مشخص است) بخش دوم برابر

خطای MSE روی دادگان تست شد با 156.4299 که همانطور هم در نمودار جعبه‌ای دیده شده بود خطای روی دادگان ارزیابی در بعد صد زیر ۲۰۰ بود است و برای دادگان تست هم به زیر ۲۰۰ آمده است



بخش سوم

در شکل زیر مشاهده میشود



سوال سوم

برای حل سوال دوم نیز از کلاس BreastCancer استفاده شده و توابع زیر هم درون آن تعریف شده:

`normalize_data`

وظیفه این تابع نرمال سازی دادگان هست

`analyst_perceptron_with_pca`

این تابع ابتدا روی دادگان روی ابعاد گفته شده `pca` زده سپس با کمک روش `cross validation` با مدل پرسپترون روی این دادگان آموزش صورت گرفته. و نتایج روش `cross validation` هم نیز در متغیر هایی ذخیره شده است.

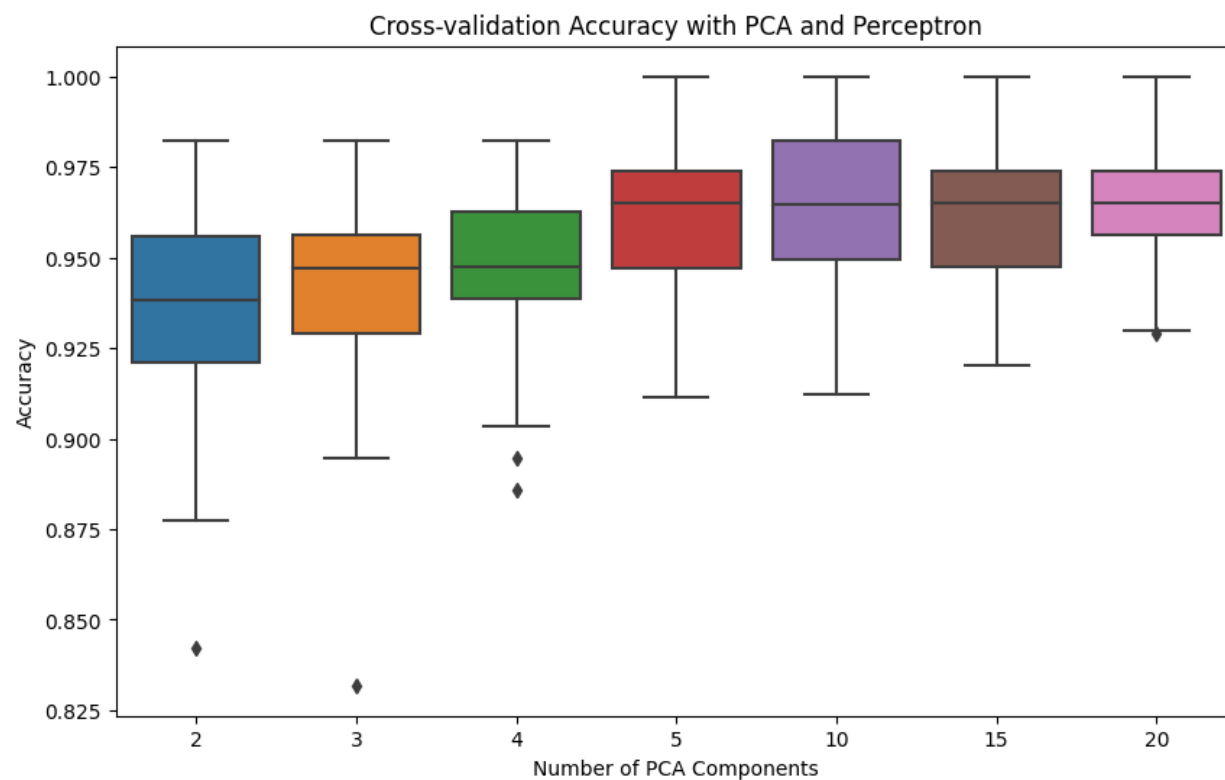
`plot_cros_validation_resualt`

این تابع نتایج حاصل از `cross validation` را بصورت نمودار جعبه‌ای نشان میدهد

در پایان هم از کلاس BreastCancer یک `object`-ای ساخته شده و ابعاد گفته شده در `pca` را تعیین کرد و در ادامه توابع مورد نیاز را فراخوانی میکنیم .

بخش سوم

همانطور که در شکل دیده میشود میانگین نمودار های جعبه‌ای در بازه‌ی 92.5 تا 95 است. همانطور که در نمودار جعبه‌ای دیده میشود کاهش بعد به بعد 5 تا حد خوبی میانگین بالایی داشته و همینطور کاهش ابعاد خوبی روی آن صورت گرفته (یعنی در بعد بیشتر از 5 میانگینی تقریباً برابر با بعد 5 داریم اما خوب تعداد ویژگی هایمان چند برابر آن است) پس کاهش بعد به بعد 5 میتواند بهترین گزینه باشد.



نتایج حاصل از Cross Validation

	2	3	4	5	10	15	20
Mean_Train_socres	0.977241	0.977241	0.977241	0.977241	0.977241	0.977241	0.977241
Mean_Test_socres	0.964673	0.964673	0.964673	0.964673	0.964673	0.964673	0.964673

کتابخانه های استفاده شده

pandas
numpy
seaborn
matplotlib
scikit-learn