CSE 151A - Homework 09

Due: Wednesday, June 10, 2020

These are extra plus problems that you can do to boost your homework score. They are entirely optional!

Plus Problem 1. (4 plus points)

Suppose we want to maximize some function $f(\vec{v})$ subject to $\|\vec{v}\| = 1$. We know that the constrained maximum is either $\vec{v}^{(1)} = (-0.28, 0.96)^\intercal$ or $\vec{v}^{(2)} = (0.6, -0.8)^\intercal$, but we don't know what the function looks like so we can't directly evaluate our function at either of these vectors. Instead, we only know that the gradient is $\nabla f(\vec{v}) = 2C\vec{v}$, where

$$C = \begin{bmatrix} 0.46 & -0.24 \\ -0.24 & 0.6 \end{bmatrix}$$

Which choice of vector corresponds to the constrained local maximum of $f(\vec{v})$? Show your work.

Solution: Recall that for a point to be at a constrained local optimum, the gradient must point in the same direction as the vector, $\nabla f(\vec{v}) = \lambda\vec{v}$.

We can compute the gradient of $f(\vec{v})$ at our two vectors.

$$\begin{aligned} \nabla f(\vec{v}^{(1)}) &= 2C\vec{v}^{(1)} \\ &= (-0.7184, 1.2864)^\intercal \end{aligned}$$

$$\begin{aligned} \nabla f(\vec{v}^{(2)}) &= 2C\vec{v}^{(2)} \\ &= (0.936, -1.248)^\intercal \end{aligned}$$

And we can perform element-wise division between each gradient and its respective vector.

$$\begin{aligned} \nabla f(\vec{v}^{(1)}) \oslash \vec{v}^{(1)} &= (-2.5657, 1.34)^\intercal \\ \nabla f(\vec{v}^{(2)}) \oslash \vec{v}^{(2)} &= (1.56, 1.56)^\intercal \end{aligned}$$

We notice that $\vec{v}^{(2)}$ is the only vector that could satisfy $\nabla f(\vec{v}) = \lambda\vec{v}$, namely when $\lambda = 1.56$. Therefore it is the only vector that could be a constrained optima. Since we are given that one of the vectors is the constrained maximum, then $\vec{v}^{(2)}$ must be the constrained maximum.

We could have alternatively looked at the Eigenvectors of C. We would observe that $\vec{v}^{(2)}$ (or $-\vec{v}^{(2)}$) is the Eigenvector of C that corresponds to the greatest Eigenvalue.

Plus Problem 2. (5 plus points)

SAT scores for six students are shown in the table below:

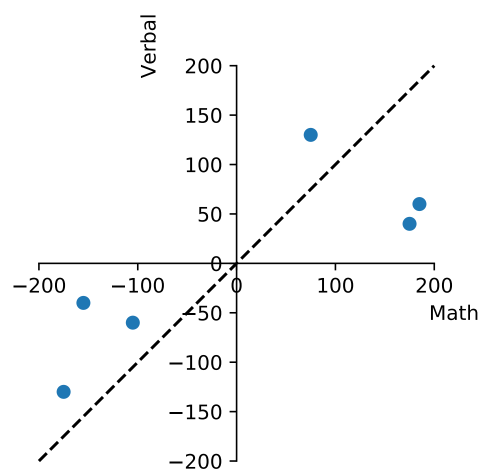
Math	Verbal
780	700
500	600
790	720
450	620
680	790
430	530

In this problem, we will use PCA to produce a single score for each student which combines their verbal and math scores.

a) Center the data by subtracting the mean math score from each math score, and the mean verbal score from each verbal score. Write your results in the form of a table like that above. The mean of each column should be zero.

> **Solution:** The average math score is 605 and the average verbal score is 660. After subtracting these from their respective columns, we get:
>
> | Math | Verbal |
> |---|---|
> | 175 | 40 |
> | -105 | -60 |
> | 185 | 60 |
> | -155 | -40 |
> | 75 | 130 |
> | -175 | -130 |

b) When plotted, the centered data looks as follows:



The dashed line is the line of 45°. Suppose $\hat{v} = (v_1, v_2)^T$ is the unit vector which lies in the direction of maximum variance. Which is larger: $|v_1|$ or $|v_2|$? Provide reasoning using the plot above.

> **Solution:** $|v_1|$ is larger. Most of the data lies below the dashed line on the right of the origin, and above the dashed line to the left of the origin. Therefore, $\hat{v}$ will point in a direction that lies closer to $(1, 0)^T$ than $(0, 1)^T$.

c) Find the $\hat{v}$ which lies in the direction of maximum variance (of the centered data). Show your work. You may use `numpy`, but you should paste your code and describe the steps taken.

> **Solution:** First we construct the matrix X whose ith column is the centered $\vec{x}^{(i)}$:
>
> ```
> >>> scores = np.array([
>         [780, 700],
> ```

```
        [500, 600],
        [790, 720],
        [450, 620],
        [680, 790],
        [430, 530]
    ])
>>> X = (scores - scores.mean(axis=0)).T
>>> X
array([[ 175., -105.,  185., -155.,   75., -175.],
       [  40.,  -60.,   60.,  -40.,  130., -130.]])
```

Then we compute the eigenvector of XX^T with the largest eigenvector:

```
>>> values, vectors = np.linalg.eigh(X @ X.T)
>>> v_hat = vectors[:,-1]
array([-0.8913125 , -0.45338949])
```

Eigenvectors are determined only up to sign. That is, $-\hat{v}$ is also an eigenvector with the same eigenvalue.

d) Use PCA to map each data vector to a single number. That is, using your solution to the previous part, compute the new variable $z^{(i)} = \vec{x}^{(i)} \cdot \hat{v}$.

Solution: The ith row of $X^T\hat{v}$ is $z^{(i)}$. Hence: `X.T @ v_hat` gives us the answer:

$z^{(i)}$
-174.28
120.76
-192.21
156.42
-125.46
214.77

Plus Problem 3. (7 plus points)

Let A be a $d \times n$ matrix and let $f : \mathbb{R}^d \to \mathbb{R}$ be the function defined by $f(\vec{x}) = \vec{x}^T AA^T \vec{x}$. Show that $\nabla f(\vec{b}) = 2AA^T\vec{b}$.

Hint: recall that $\nabla f(\vec{b})$ denotes the gradient of f at $\vec{b}$. Is is defined to be:

$$\nabla f(\vec{b}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{b}) \\ \frac{\partial f}{\partial x_2}(\vec{b}) \\ \vdots \\ \frac{\partial f}{\partial x_d}(\vec{b}) \end{pmatrix}$$

Solution: We start by computing the partial derivative of f with respect to x_1. We have:

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_1}(\vec{x}^T AA^T \vec{x})$$

We are computing the derivative with respect to x_1, but x_1 doesn't appear explicitly in the above; it is hidden within $\vec{x}$. To find x_1, first recall from lecture that $\vec{x}^T AA^T \vec{x} = \sum_{i=1}^n (\vec{a}^{(i)} \cdot \vec{x})^2$, where $\vec{a}^{(i)}$ is the ith column of A. Therefore:

$$= \frac{\partial f}{\partial x_1} \left(\sum_{i=1}^n (\vec{a}^{(i)} \cdot \vec{x})^2 \right)$$

We can push the derivative inside of the summation:

$$= \sum_{i=1}^n \frac{\partial f}{\partial x_1} (\vec{a}^{(i)} \cdot \vec{x})^2$$

Using the chain rule:

$$= \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{x}) \cdot \frac{\partial f}{\partial x_1} (\vec{a}^{(i)} \cdot \vec{x})$$

x_1 is still hidden inside of $\vec{x}$, but we can expose it by expanding the dot product:

$$\begin{aligned}
&= \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{x}) \cdot \frac{\partial f}{\partial x_1} ({a^{(i)}}_1 x_1 + {a^{(i)}}_2 x_2 + \ldots + {a^{(i)}}_d x_d) \\
&= \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{x}) \cdot ({a^{(i)}}_1 + 0 + \ldots + 0) \\
&= \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{x}) {a^{(i)}}_1
\end{aligned}$$

This is $\frac{\partial f}{\partial x_1}$. Calculating the derivative with respect to x_2 is exactly the same as the above, except ${a^{(i)}}_1$ is replaced by ${a^{(i)}}_2$:

$$\frac{\partial f}{\partial x_2} = \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{x}) {a^{(i)}}_2$$

And so the gradient vector is:

$$\nabla f(\vec{b}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{b}) \\ \frac{\partial f}{\partial x_2}(\vec{b}) \\ \vdots \\ \frac{\partial f}{\partial x_d}(\vec{b}) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{b}) {a^{(i)}}_1 \\ \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{b}) {a^{(i)}}_2 \\ \vdots \\ \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{b}) {a^{(i)}}_d \end{pmatrix}$$

Working from the other direction, we now want to show that $2AA^T\vec{b}$ is equal to this vector. We have:

$$2AA^T\vec{b} = 2 \begin{pmatrix} \vec{a}^{(1)} & \vec{a}^{(2)} & \cdots & \vec{a}^{(n)} \\ \Bigg\downarrow & \Bigg\downarrow & \cdots & \Bigg\downarrow \end{pmatrix} \begin{pmatrix} \vec{a}^{(1)} & \longrightarrow \\ \vec{a}^{(2)} & \longrightarrow \\ \vdots & \vdots \\ \vec{a}^{(n)} & \longrightarrow \end{pmatrix} \vec{b}$$

The product of $A^T\vec{b}$ will be a vector whose ith entry is given by the dot product of the ith row of A^T with $\vec{b}$. The ith row of A^T is simply the ith column of A; that is, it is $\vec{a}^{(i)}$:

$$
\begin{aligned}
&= 2 \begin{pmatrix} \vec{a}^{(1)} & \vec{a}^{(2)} & \cdots & \vec{a}^{(n)} \\ \Bigg\downarrow & \Bigg\downarrow & \cdots & \Bigg\downarrow \end{pmatrix} \begin{pmatrix} \vec{a}^{(1)} \cdot \vec{b} \\ \vec{a}^{(2)} \cdot \vec{b} \\ \vdots \\ \vec{a}^{(n)} \cdot \vec{b} \end{pmatrix} \\
&= 2 \left[(\vec{a}^{(1)} \cdot \vec{b})\vec{a}^{(1)} + (\vec{a}^{(2)} \cdot \vec{b})\vec{a}^{(2)} + \ldots + (\vec{a}^{(n)} \cdot \vec{b})\vec{a}^{(n)} \right] \\
&= 2 \sum_{i=1}^{n} (\vec{a}^{(i)} \cdot \vec{b})\vec{a}^{(i)} \\
&= \begin{pmatrix} \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{b}) {a^{(i)}}_1 \\ \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{b}) {a^{(i)}}_2 \\ \vdots \\ \sum_{i=1}^n 2(\vec{a}^{(i)} \cdot \vec{b}) {a^{(i)}}_d \end{pmatrix}
\end{aligned}
$$

This is exactly what we found for $\nabla f(\vec{b})$, and so $\nabla f(\vec{b}) = 2AA^T\vec{b}$, as expected.