PCA (Lagrange multiplier) and SVD

PCA motivation

To identify the subspace in which the data approximately lies.

Assumption of PCA

the desired subspace data approximation is the one that has maximized variance when data is projected to that space.

Target:

To find the **unit vector** u so that then the data is **projected** into the direction corresponding to u, the variance of the projected data is maximized.

How to process PCA?

1. normalized each feature to have mean 0 and variance 1.

$$x_j^{(i)} \leftarrow \frac{x_j^{(i)} - \mu_j}{\sigma_j}$$

2. calculated the variance of variance

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}({x^{(i)}}^T u)^2 &= \frac{1}{n}\sum_{i=1}^{n} u^T x^{(i)} {x^{(i)}}^T u \\ &= u^T\left(\frac{1}{n}\sum_{i=1}^{n} x^{(i)} {x^{(i)}}^T\right) u.\end{aligned}$$

why xTu is the variance>

xTu = |u|*|x| cos (theta)

since |u| = 1

xTu = |x| cos the length of projection of x onto u

why xxT is covariance matrix?

$$\begin{aligned}V = \frac{1}{n}\sum_{i=1}^{n} y_i^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^\top \mathbf{v}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}\mathbf{x}_i^\top \mathbf{v}\cdot\mathbf{x}_i^\top \mathbf{v} = \frac{1}{n}\sum_{i=1}\mathbf{v}^\top \mathbf{x}_i \cdot \mathbf{x}_i^\top \mathbf{v} \\ &= \mathbf{v}^\top \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\right)}_{\text{Covariance matrix}} \mathbf{v} = \mathbf{v}^\top C \mathbf{v}\end{aligned}$$

Mind that the covariance matrix is defined as:

$$C = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i - \mu\right)\left(\mathbf{x}_i - \mu\right)^\top$$

Where μ is the mean of $\mathbf{x}_i, i = 1, 2, \ldots, n$. Which is why we asked for the variables $\mathbf{x}_i$ to be centered. So that $\mu = 0$ and the formula for covariance would simplify to $C = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top$.

Turned out the **definition** of the **covariance matrix** is actually the sum of |xi - nu|^2

But we didn;t calculate mu. **how our variance calculation end up with covariance matrix?**

A: only after centering that make mu = 0, the result of variance calculation can meet the definition of the **covariance matrix**.

3. maximize the variance

is the same as getting this principal eigenvector of sigma

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} x^{(i)} x^{(i)T},$$

find maxima with Lagrange multiplier

https://ekamperi.github.io/mathematics/2020/11/01/principal-component-analysis-lagrange-multiplier.html

$$L(x,\lambda) = f(x) - \lambda(g(x) - c)$$

$$f(x) = V = v^T C v$$

$$g(x) = v^T v - 1 = 0$$

$$L(v,\lambda) = v^T C v - \lambda(v^T v - 1). \qquad C = \frac{1}{n}\sum_{i=1}^{n} x x^T \text{ (symmetric)}$$

$$\left\{\begin{aligned} \frac{\partial L}{\partial v} &= 2v^T C - 2v^T \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= v^T v - 1 = 0 \end{aligned}\right.$$

$$Cv = \lambda v$$

$$\boxed{\begin{aligned} &\alpha = x^T A x \\ &\frac{\partial \alpha}{\partial x} = x^T(A + A^T) \\ &\text{when } A \text{ is symmetric} \\ &A = A^T \\ &\frac{\partial \alpha}{\partial x} = 2x^T A. \end{aligned}}$$

understanding of the last equation (why lam as the parameter of Lagrange could be interpreted as eigenvalue):

Same as Ax = lamx, Here we get Cv = lamv.

Though lam comes from Lagrange equation, but it meet the eigenvector equation. Thus the resolution of v comes from the eigenvectors covariance matrix C

The connection of PCA and SVD:

From the conclusion above, PCA process is to find the eigenvector of covariance matrix C, which is XX^T. That is exactly the (definition in SVD of)U for x.

Understand of Lagrange multiplier:

The motivation of Lagrange multiplier

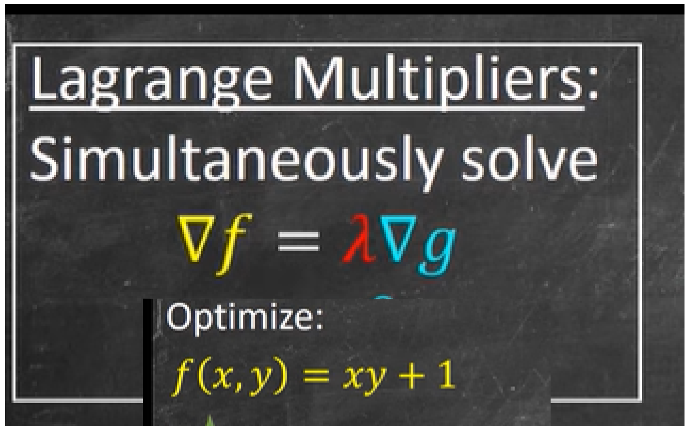
The Lagrange multiplier is a strategy for finding the local maxima and minima of a function subject to equality constraints.

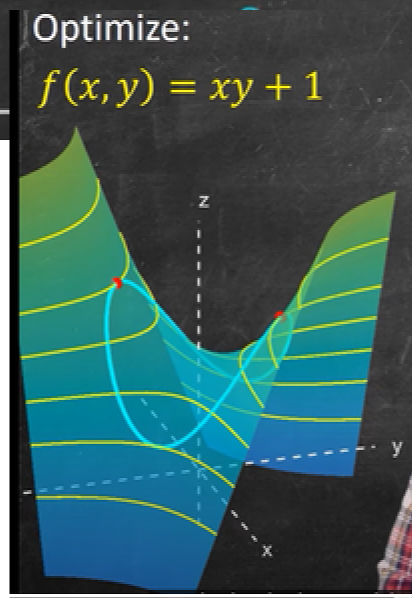
What is the Lagrange multiplier?

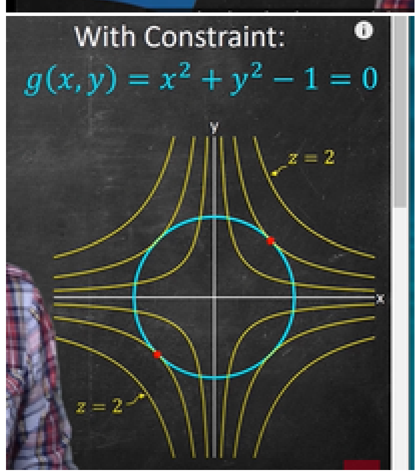
The method can be summarized as follows: in order to find the maximum or minimum of a function $f(x)$ subjected to the equality constraint $g(x) = 0$, form the Lagrangian function

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$$

Geometric Meaning:







if constraint qualification applies, then the gradient of the constraint function can be expressed as a linear combination of the gradients of the constraints, with the Lagrange multipliers acting as coefficients

Example of Lagrange multiplier

Lagrange Multipliers:

Simultaneously solve

$$\nabla f = \lambda \nabla g$$
$$g = 0$$

$$f(x, y) = xy + 1 \Rightarrow \nabla f = \langle y, x \rangle$$

$$g(x, y) = x^2 + y^2 - 1 \Rightarrow \nabla g = \langle 2x, 2y \rangle$$

$$\boxed{\begin{aligned} y &= \lambda 2x \\ x &= \lambda 2y \\ x^2 + y^2 - 1 &= 0 \end{aligned}}$$