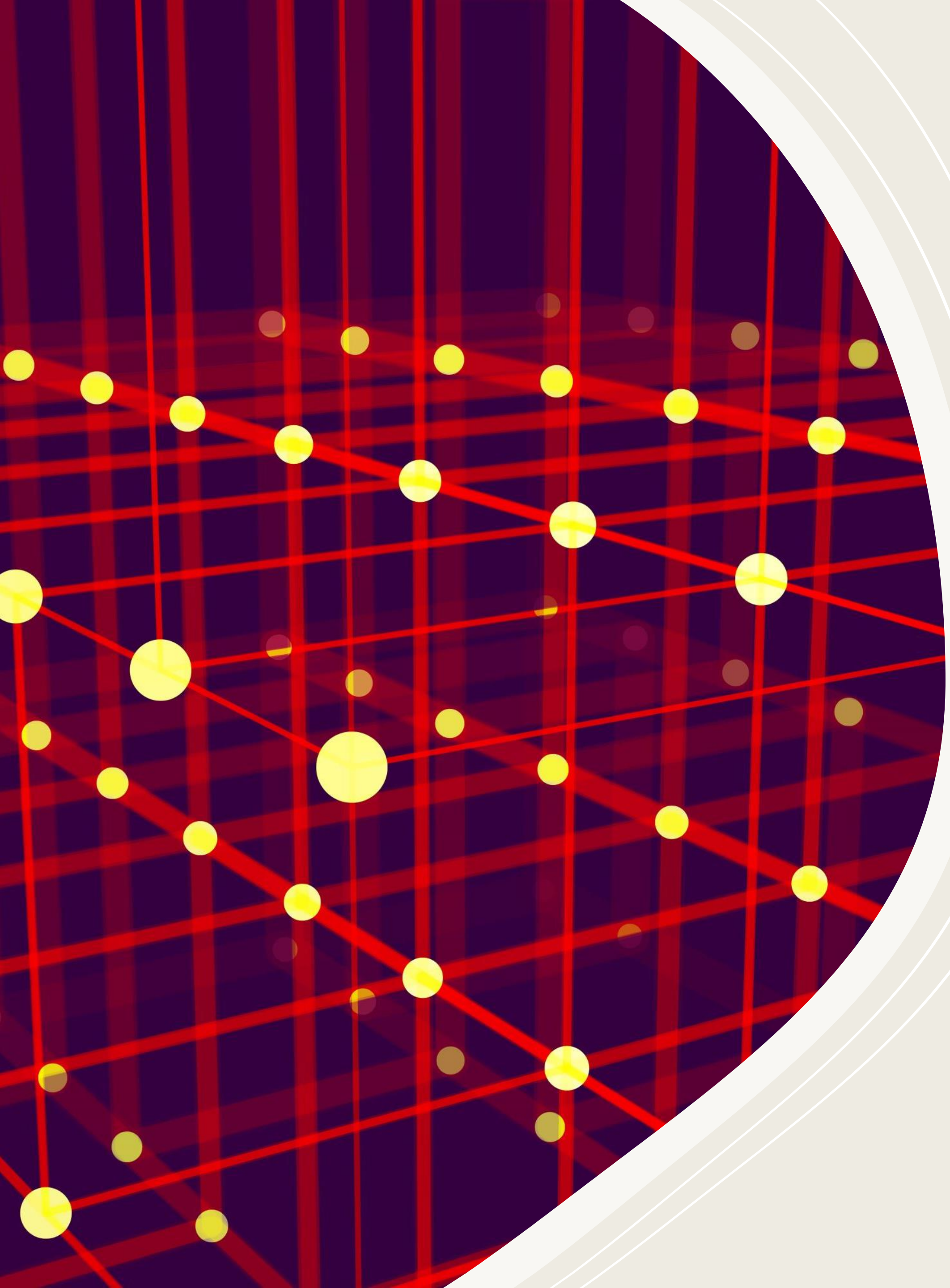

CPSC 340 – Tutorial 9

Lironne Kurzman
lironnek@cs.ubc.ca

Slides courtesy of Nam Hee Kim

University of British Columbia

November 22nd, 2021

Agenda

1. Robust PCA
2. Gradient for Robust PCA

Robust PCA

Very similar objective function but we use L1 distance rather than L2

$$f(Z,W) = \sum_{i=1}^{n} \sum_{j=1}^{d} |\langle w^j, z_i \rangle - x_{ij}|,$$

Like the PCA you've seen before we need to optimize W and Z!

Approximating Robust PCA

Very similar objective function but we use L1 distance rather than L2

$$f(Z, W) = \sum_{i=1}^{n} \sum_{j=1}^{d} |\langle w^j, z_i \rangle - x_{ij}|,$$

But we want a smooth approximation using:

$$|\alpha| \approx \sqrt{\alpha^2 + \epsilon},$$

Evaluating Robust PCA

Given the definition of f and the smooth approximation we have to compute f

$$f(Z,W) = \sum_{i=1}^{n} \sum_{j=1}^{d} |\langle w^j, z_i \rangle - x_{ij}|, \qquad |\alpha| \approx \sqrt{\alpha^2 + \epsilon},$$

Putting these together we get:

$$f(Z,W) = \sum_{i=1}^{n} \sum_{j=1}^{d} \sqrt{(\langle w^j, z_i \rangle - x_{ij})^2 + \epsilon}$$

Evaluating Robust PCA

Given the definition of f and the smooth approximation we have to compute f

$$f(Z,W) = \sum_{i=1}^{n}\sum_{j=1}^{d} |\langle w^j, z_i\rangle - x_{ij}|, \qquad |\alpha| \approx \sqrt{\alpha^2 + \epsilon},$$

(SLIDE 6 of Lecture “More PCA”)

- The 3 different ways to write the PCA objective function:

$$f(W,Z) = \sum_{i=1}^{n}\sum_{j=1}^{d} (\langle w^j, z_i\rangle - x_{ij})^2 \quad \text{(approximating } x_{ij} \text{ by } \langle w^j, z_i\rangle)$$

$$= \sum_{i=1}^{n} \|W^T z_i - x_i\|^2 \quad \text{(approximating } x_i \text{ by } W^T z_i)$$

$$= \|ZW - X\|_F^2 \quad \text{(approximating } X \text{ by } ZW)$$

Evaluating Robust PCA

Given the definition of f and the smooth approximation we have to compute f

$$f(Z,W) = \sum_{i=1}^{n} \sum_{j=1}^{d} |\langle w^j, z_i \rangle - x_{ij}|, \qquad |\alpha| \approx \sqrt{\alpha^2 + \epsilon},$$

Putting these together we get (with some abuse of notation):

$$f(Z,W) = \sum \sqrt{(ZW - X)^2 + \epsilon}$$

Gradient of Robust PCA

Combining the equations we already have:

$$f(Z, W) = \sum_{i=1}^{n} \sum_{j=1}^{d} |\langle w^j, z_i \rangle - x_{ij}|, \qquad |\alpha| \approx \sqrt{\alpha^2 + \epsilon},$$

We get:

$$f(Z, W) = \sum_{i=1}^{n} \sum_{j=1}^{d} \sqrt{(\alpha)^2 + \epsilon}$$

$$\alpha = \langle w^j, z_i \rangle - x_{ij}$$

Gradient of Robust PCA

If we were to take the gradient w.r.t w^j

$$f(Z, W) = \sum_{i=1}^{n} \sum_{j=1}^{d} \sqrt{(\alpha)^2 + \epsilon} \qquad \boxed{\alpha = \langle w^j, z_i \rangle - x_{ij}}$$

We get:

$$\frac{df(\alpha)}{dw^j} = \sum_{i=1}^{n} \frac{1}{2}(\alpha^2 + \epsilon)^{-\frac{1}{2}} \cdot 2\alpha \cdot \frac{d\alpha}{dw^j}$$

$$\frac{df(\alpha)}{dw^j} = \sum_{i=1}^{n} \frac{\alpha}{\sqrt{\alpha^2 + \epsilon}} \cdot \frac{d\alpha}{dw^j}$$

What is da/dw^j ?

Gradient of Robust PCA

If we were to take the gradient w.r.t W_j

$$f(Z, W) = \sum_{i=1}^{n} \sum_{j=1}^{d} \sqrt{(\alpha)^2 + \epsilon} \qquad \alpha = \langle w^j, z_i \rangle - x_{ij}$$

We get:

$$\boxed{\frac{df(\alpha)}{dw^j}} = \sum_{i=1}^{n} \frac{1}{2}(\alpha^2 + \epsilon)^{-\frac{1}{2}} \cdot 2\alpha \cdot \frac{d\alpha}{dw^j}$$

$$\frac{df(\alpha)}{dw^j} = \sum_{i=1}^{n} \frac{\alpha}{\sqrt{\alpha^2 + \epsilon}} \cdot \frac{d\alpha}{dw^j}$$

How does this formula change when we take the derivative w.r.t z_i?

Gradient of Robust PCA

Again we can see that this scenario is similar to what we've seen in class before

HINT (slide 36 of More PCA lecture)

PCA Computation: Alternating Minimization

- With centered data, the PCA objective is:

$$f(W,Z) = \sum_{i=1}^{n} \sum_{j=1}^{d} (\langle w^j, z_i \rangle - x_{ij})^2$$

$$\nabla_W f(W,Z) = Z^T Z W - Z^T X$$

(writing gradient as a matrix)

$$\nabla_Z f(W,Z) = Z W W^T - X W^T$$

Gradient of Robust PCA

Again we can see that this scenario is similar to what we've seen in class before

HINT (slide 36 of More PCA lecture)

PCA Computation: Alternating Minimization

- With centered data, the PCA objective is:

$$f(W,Z) = \sum_{i=1}^{n} \sum_{j=1}^{d} (\langle w^j, z_i \rangle - x_{ij})^2$$

$$\nabla_W f(W,Z) = Z^T Z W - Z^T X$$

(writing gradient as a matrix)

$$\nabla_Z f(W,Z) = Z W W^T - X W^T$$

Demo

Eigenfaces