CS181: Principal Components Analysis (PCA)

	non prob	prob
discrete z	✓	✓
cont. z	TODAY	

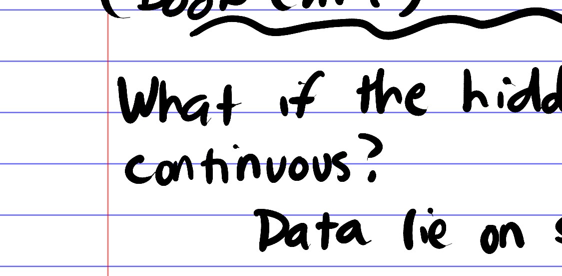
unsupervised

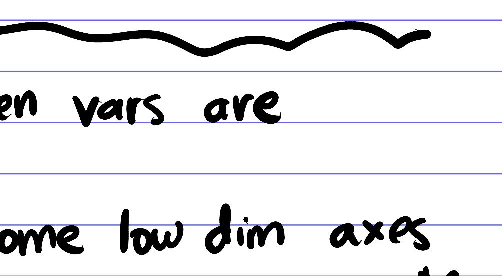
Notes:
- HW5 out, practical out Friday
- After "CUBE": structured models, RL

(Book Ch. 7)

What if the hidden vars are continuous?

Data lie on some low dim axes





Today: start w/ linear version of identifying these "hidden axes"

$\underline{x}$ is a D-dim vec.

goal: rep $\underline{x} \approx z_1 \underline{U}_1 + z_2 \underline{U}_2 + \dots z_K \underline{U}_K$

scalars that rep. location of x is basis U

$\underline{U}$ are D-dim basis vecs.

s.t. $K < D$

$$\mathcal{L}(\{\underline{z}_n\}, U) = \frac{1}{N}\sum_n \| \underbrace{\underline{x}_n}_{\text{orig } x} - \underbrace{U \underline{z}_n}_{\text{recon } x} \|_2^2$$

$D[\;]$ $\quad D[\overset{K}{\;}][K]$

Is the solution to the above loss unique?

Let's imagine: $U' = UQ$

$\underline{z}_n' = Q^{-1} \underline{z}_n$

$U' \underline{z}_n' = UQQ^{-1}\underline{z}_n = U\underline{z}_n$

No! Not unique!

Lots of works solve for U w/ diff contraints: U sparse, U non-neg.

Today: U is orthonormal

$\langle U_k, U_k \rangle = 1$

$\langle U_k, U_{k'} \rangle = 0$

} this will correspond to PCA

- Nice linear alg properties:

$x \approx z_1 U_1 + \dots z_k U_k + \dots$ all elements are $\perp$

$U_k^T x \quad z_1 \underbrace{U_k^T U_1}_{0} + \dots z_k \underbrace{U_k^T U_k}_{1} + \dots \underbrace{\varnothing}_{0} \; \underbrace{\varnothing}_{0}$

Given $U, \underline{x}$ easy to find coordinates in U-space for x_n, the $\underline{z}_n$

- Let's make this more interpretable by subtracting $\bar{x}$, mean of data:

$$\underline{x} = \underbrace{\bar{\underline{x}}}_{\text{mean of the data}} + \underbrace{z_1 \underline{U}_1 + z_2 \underline{U}_2 + \dots z_K \underline{U}_K}_{\text{bases capture variation from mean}}$$

- Back to our loss:

$$\mathcal{L}(\{z\}, U) = \frac{1}{N}\sum_n \| (\underline{x}_n - \underline{\bar{x}}) - U \underline{z}_n \|_2^2$$

($\underline{\bar{x}}$: fixed)

constraint: U orthonormal

To solve: Note from linear algebra: if U was $D \times D$ then recon would be perfect.

$x_n = \bar{x} + z U_1 + \dots z_D U_D$

Let's substitute:

$$\mathcal{L} = \frac{1}{N}\sum_n \| \sum_{k=1}^{D} z_{nk} U_k - \sum_{k=1}^{K} z_{nk} U_k \|_2^2$$

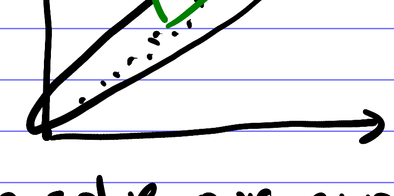
$$\mathcal{L} = \frac{1}{N}\sum_n \| \sum_{k=K+1}^{D} z_{nk} U_k \|_2^2$$

note: $\langle U_k, U_k \rangle = 1$, $\langle U_k, U_{k'} \rangle = 0$

$$= \frac{1}{N}\sum_n \sum_{k=K+1}^{D} z_{nk}^2$$

$$= \frac{1}{N}\sum_n \sum_{k=K+1}^{D} \underbrace{U_k^T (\underline{x}_n - \underline{\bar{x}})}_{z_{nk}} \underbrace{(\underline{x}_n - \underline{\bar{x}})^T U_k}_{z_{nk}}$$

$$= \sum_{k=K+1}^{D} U_k^T \underbrace{\left[\frac{1}{N}\sum_n (\underline{x}_n - \underline{\bar{x}})(\underline{x}_n - \underline{\bar{x}})^T \right]}_{\text{this is the (empirical) covariance of } x.\ \Sigma} U_k$$



To solve our expression above: suppose we only had one vec $\underline{u}$ to optimize

$$\min_{\underline{u}} \underline{u}^T \Sigma \underline{u} \quad \text{s.t.} \quad \underline{u}^T \underline{u} = 1$$

(put constraint into obj w/ Lagrange multiplier:

$$\min_{\underline{u}} \underline{u}^T \Sigma \underline{u} - \lambda u^T u$$

↳ take grad of this obj w.r.t. $\underline{u}$; set $= 0$

$$\Sigma \underline{u} = \lambda \underline{u}$$

eigenvalue problem!!

we would $\underline{u}$ to be eigenvec. w/ smallest eigenval λ as the solution

Back to the main problem:

$$\boxed{\min_{U} \sum_{k=K+1}^{D} U_k^T \Sigma U_k \quad \text{s.t. } U_k\text{'s are orthonormal eigenvectors}}$$

solution: correspond to leaving out eigenvectors w/ the smallest $K+1 \dots D$ eigenvalues.

alternatively: keep U_k [eigenvecs of Σ) w/ the K biggest eigenvalues.

(Hooray! Just an SVD!!)

Notes:

1) We motivated this prob by minimizing recon error (goal: z, U s.t. $\underline{x} \approx U^T \underline{z}$)

"Reconstruction View" of PCA

2) Alternate view:

"Variance Preservation View"

find the z, U s.t. most of the variance in x is captured

$z = \underline{U}^T \underline{x}$, z is one-dim

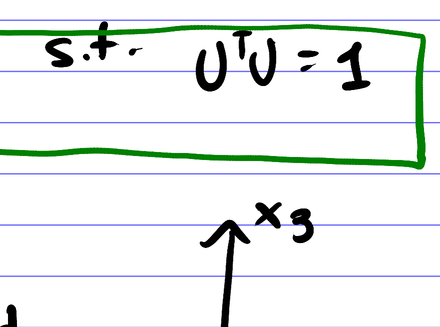
$$\text{var}(z) = U^T \underbrace{\text{var}(x)}_{\Sigma} U$$

goal: $\boxed{\max_{U} U^T \Sigma U \quad \text{s.t.} \quad U^T U = 1}$

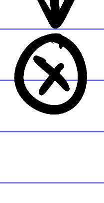
3) Uniqueness:

subspace found by PCA is unique

BUT: recon view prob has other soln s (orthonormal) that are equally good.



Last note: Prelude to next time:



$z \sim N(0, 1)$

$x = U^T z + \epsilon$

}
- Prob. PCA
- Factor Analysis
- Indep. Component Analysis
- Autoencoders/VAE