

Data visualization

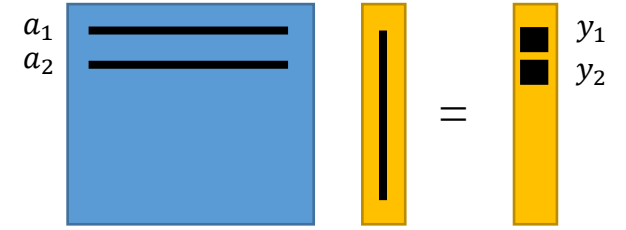
Q&A 7

David Hoksza

<http://siret.ms.mff.cuni.cz/hoksza>

PCA

- Proč je u $y = A^T(x - \mu)$ matice vlastních vektorů transponovaná
- Nejvíc používaný score plot. Kdy ostatní
 - Interpretovatelnost
- Normalizace
 - Proč normalizovat?
 - Lze dělat PCA pro nenormalizovaná data?
 - Normlizovaná data – znamená to, že mají stejné hodnoty?
- Loading, scores
 - Jsou loadings vlastní vektory kovarianční matice? - ANO
 - Score by mělo odpovídat vlastním vektorům (eigenvectors) a loading pak odpovída vlastním cislum (eigenvalues) covariacni matice - NE

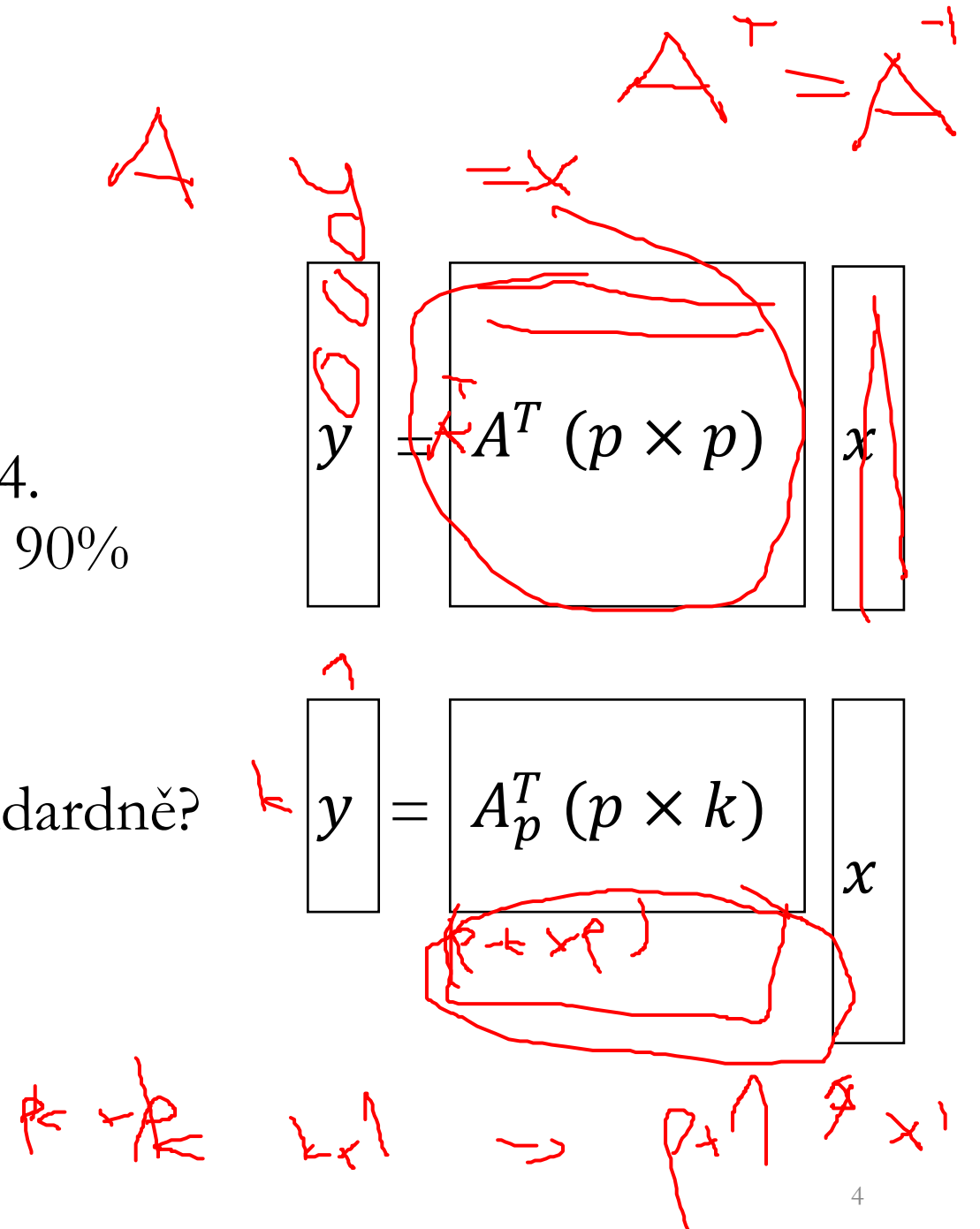


PCA

- Je účel PCA redukovat dimenzi na něco rozumného, co je vizualizovatelné, aby se uchovala nejdůležitější informace?
- Má smysl využít data, která jsou nejméně informativní?
- Rozlišují se překrývající se body?

Příklady

- Konkrétní příklad použití Iris datasetu
- U Iris datasetu když převádím do 3D, tak 4. sloupec ztratím, což nevadí, protože mám 90% variability?
- Používá se PCA na grayscale obrázky standardně?
- Jak se u grayscale obrázků dostat zpět do původního prostoru?



Ostatní techniky

- Co když není cumulative proportion na prvních třech komponentách dost vysoká a potřebuji vizualizovat? Lze to obejít?
- Jaké jsou jiné alternativy k PCA a kdy se hodí více?
- Používá se PCA i k data miningu ve smyslu pochopení dat?

Ostatní techniky

- Použít deep learning + např. cosine similarity
- PCA vs LDA
 - LDA supervised (maximization of class separability)
 - Co je lepší pro rozpoznání obličeje?

