TAM 598 Lecture 18 :

Unsupervised Learning – Dimensionality Reduction

Announcements:

- No class on Wednesday
- HW 5 covers lectures 17-20; due on Fri Apr 18

Dimensionality Reduction -

- we have observations $\underset{\sim}{x}_{1:n}$, each of which is a high-dimensional vector $\underset{\sim}{x}_i \in \mathbb{R}^D$, with $D \gg 1$.
- our goal is to describe the data set with a smaller number of dimensions without losing too much information, ie to project each $\underset{\sim}{x}_i$ to a d-dimensional vector $\underset{\sim}{z}_i$ where $d \ll D$.

why? for visualization
to do clustering, density estimation, ...
for supervised learning tasks

Principal Component Analysis :

-

-

-

To find the matrices $\underline{\underline{W}}$, $\underline{\underline{V}}$ we minimize the reconstruction error.

$$L(\underline{\underline{W}}, \underline{\underline{V}}, \underline{x}_0) =$$

Taking $\frac{\partial L}{\partial \underline{\underline{W}}} = 0$, $\frac{\partial L}{\partial \underline{\underline{V}}} = 0$ and solving:

Specifically:

- let $\underset{\sim}{u}_i$ and λ_i be the i^{th} eigenvalue of $\underline{\underline{C}}$, sorted so

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$$

- then

- our projection map becomes

- the reconstruction map is

* The sum of the first k eigenvectors

$$\sum_{i=1}^{k} \lambda_i$$

* the minimum reconstruction error is

$$L(\underline{\underline{W}}, \underline{\underline{V}}, \underset{\sim}{x}_0) = \frac{1}{n} \sum_{i=1}^{n} \| \underset{\sim}{x}_i - g(f(\underset{\sim}{x}_i)) \|^2$$

$$= \sum_{j=d+1}^{D} \lambda_j$$

probabilistic interpretation: assume our data points $\underset{\sim}{x}_{1:n}$ are generated by a linear Gaussian model

and latent variables $\underset{\sim}{z}_i$ are generated by a Gaussian prior

We can maximize the marginal likelihood:

which is a gaussian. Maximizing $\log p(\underset{\sim}{x}_{1:n})$ gives the same result as before, now with variance