Principal Component Analysis

Principal Component Analysis

E-step

$$q(t_i) = p(t_i \mid x_i, \theta) = \frac{p(x_i \mid t_i, \theta) p(t_i)}{Z}$$

$$= \mathcal{N}(\widetilde{\mu}_i, \widetilde{\Sigma}_i)$$

Principal Component Analysis

M-step

$$\max_{\theta} . \mathbb{E}_{q(T)} \sum_i \log p(x_i \mid t_i, \theta) p(t_i)$$

$$= \sum_i \mathbb{E}_{q(t_i)} \log \left(\frac{1}{Z} \exp\left(\ldots \right) \exp\left(\ldots \right) \right)$$

(as defined in previous slide - likelihood)

Principal Component Analysis

M-step

$$\max_{\theta}. \mathbb{E}_{q(T)} \sum_i \log p(x_i \mid t_i, \theta) p(t_i)$$

$$= \sum_i \log \frac{1}{Z}$$

Z doesn't depend on theta. So we can just separate it as follows:

$$+ \sum_i \mathbb{E}_{q(t_i)} \log \left(\exp\left(\ldots\right) \exp\left(\ldots\right) \right)$$

Principal Component Analysis

M-step

$$\max_{\theta}.\ \mathbb{E}_{q(T)} \sum_i \log p(x_i \mid t_i, \theta) p(t_i)$$

$$= \sum_i \log \frac{1}{Z}$$

$$+ \sum_i \mathbb{E}_{q(t_i)} \log \left(\exp\left(\ldots\right) \exp\left(-\frac{t_i^2}{2} \right) \right)$$

in the one dimensional case. This is just standard normal pdf.

Principal Component Analysis

This is the squared distance from the actual point x_i and the projection of t_i into the multidimensional space as a linear function Wt_i + b.
Smart!
We defined it as the probability that x_i is equal to something given the parameters theta (I believe W and b are included in this theta), and latent variable t_i. If the probability is high, that means the distance shouldn't be that far. They gotta be close to each other.

M-step

$$\max_{\theta} . \mathbb{E}_{q(T)} \sum_i \log p(x_i \mid t_i, \theta) p(t_i)$$

$$= \sum_i \log \frac{1}{Z}$$

$$+ \sum_i \mathbb{E}_{q(t_i)} \underbrace{\log \left(\exp\left(-\frac{(x - W t_i - b)^2}{2\sigma^2} \right) \exp\left(-\frac{t_i^2}{2} \right) \right)}_{a t_i^2 + c t_i + d}$$

And this is becauase we take the log of the exp. Aha! ^^

And we can compute this analytically because it is quadratic. Simple equation.

Summary

Probabilistic formulation of PCA

- Allows for missing values
- Straightforward iterative scheme for large dimensionalities
- Can do mixture of PPCA
- Hyperparameter tuning (number of components or choose between diagonal and full covariance)