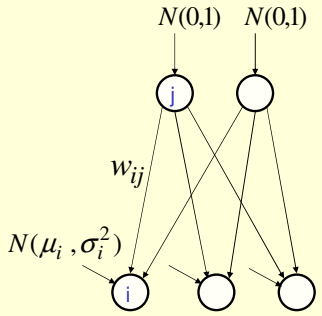
# CSC2515:

## Lecture 7 (prelude)
## Some linear generative models and a coding perspective

Geoffrey Hinton

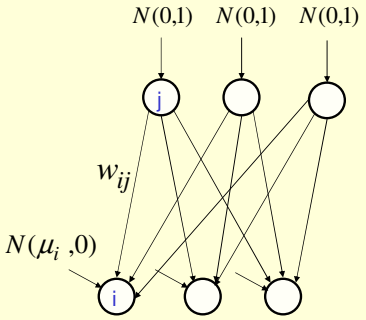
## The Factor Analysis Model

- The generative model for factor analysis assumes that the data was produced in three stages:
  - Pick values independently for some hidden factors that have Gaussian priors
  - Linearly combine the factors using a factor loading matrix. Use more linear combinations than factors.
  - Add Gaussian noise that is different for each input.



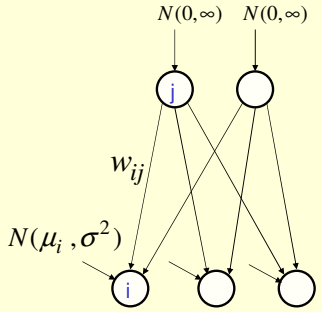

## The Full Gaussian Model

- The generative model for factor analysis assumes that the data was produced in three stages:
  - Pick values independently for some hidden factors that have Gaussian priors
  - Linearly combine the factors using a square matrix.
  - There is no need to add Gaussian noise because we can already generate all points in the dataspace.



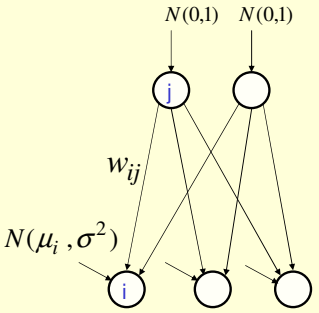

## The PCA Model

- The generative model for factor analysis assumes that the data was produced in three stages:
  - Pick values independently for some hidden factors that can have any value
  - Linearly combine the factors using a factor loading matrix. Use more linear combinations than factors.
  - Add Gaussian noise that is the same for each input.



## The Probabilistic PCA Model

- The generative model for factor analysis assumes that the data was produced in three stages:
  - Pick values independently for some hidden factors that can have any value
  - Linearly combine the factors using a factor loading matrix. Use more linear combinations than factors.
  - Add Gaussian noise that is the same for each input.




## A coding view of FA, PPCA and PCA

- Factor analysis pays to communicate the hidden factor values:
  - log p(value|gaussian)
- It also pays to communicate the residual errors in each observed value:
  - log p(residual|noise model for that dimension)
- PPCA pays both costs but uses the same noise model for all data dimensions (suboptimal)
- PCA ignores the cost of communicating the factor values. It also uses the same noise model for all input dimensions.

## A big difference in behaviour of FA and PCA

- Suppose we have data in which dimensions A and B have very small variance but very high correlation and dimension C has high variance but no correlation with the other dimensions.
- With only one factor, factor analysis will choose to represent what is common to A and B.
  - It wouldn't save anything by representing C as with its factor because it still has to communicate it under a Gaussian.
- With only one factor, PCA will represent C.
  - It can send the factor value for free.