

Latent Variable Models & Expectation Maximization

Week 2

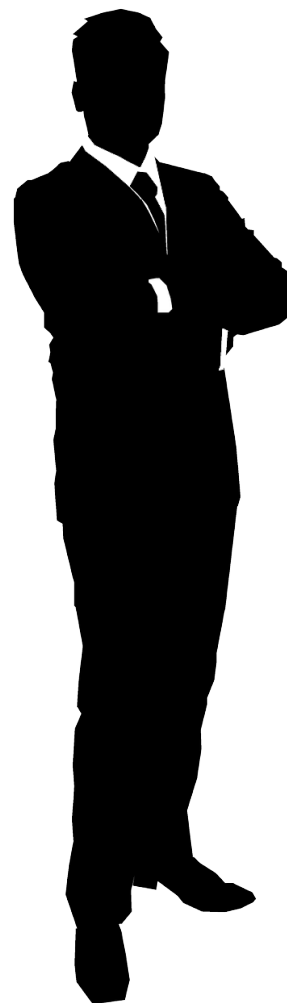
- What is a latent variable, why do we need it, and how to use it
- Common latent variable models (clustering and dimensionality reduction)

Week 2

- What is a latent variable, why do we need it, and how to use it
- Common latent variable models (clustering and dimensionality reduction)
- How to train them with Expectation Maximization algorithm
- Extensions of Expectation Maximization such as handling missing data

**Latent (hidden) variable is a variable
that you never observe**







	High school grade
<i>John</i>	4.0
<i>Helen</i>	3.7
<i>Jack</i>	3.2
<i>Emma</i>	2.9

	High school grade	University grade
<i>John</i>	4.0	4.0
<i>Helen</i>	3.7	3.6
<i>Jack</i>	3.2	N/A
<i>Emma</i>	2.9	3.2

	High school grade	University grade	IQ score
<i>John</i>	4.0	4.0	120
<i>Helen</i>	3.7	3.6	N/A
<i>Jack</i>	3.2	N/A	112
<i>Emma</i>	2.9	3.2	N/A

	High school grade	University grade	IQ score	Phone Interview
<i>John</i>	4.0	4.0	120	3/4
<i>Helen</i>	3.7	3.6	N/A	4/4
<i>Jack</i>	3.2	N/A	112	2/4
<i>Emma</i>	2.9	3.2	N/A	3/4

	High school grade	University grade	IQ score	Phone Interview	Onsite interview
<i>John</i>	4.0	4.0	120	3/4	?
<i>Helen</i>	3.7	3.6	N/A	4/4	?
<i>Jack</i>	3.2	N/A	112	2/4	?
<i>Emma</i>	2.9	3.2	N/A	3/4	?

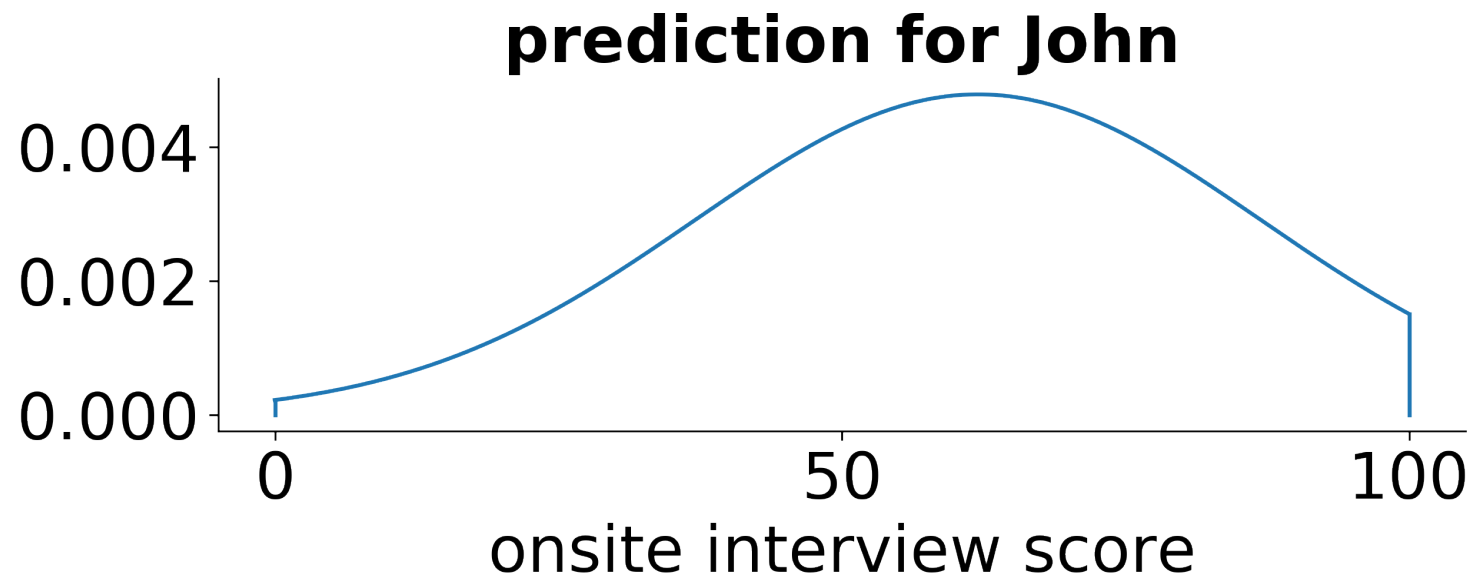
	High school grade	University grade	IQ score	Phone Interview	Onsite interview
<i>John</i>	4.0	4.0	120	3/4	?
<i>Helen</i>	3.7	3.6	N/A	4/4	?
<i>Jack</i>	3.2	N/A	112	2/4	?
<i>Emma</i>	2.9	3.2	N/A	3/4	?

	High school grade	University grade	IQ score	Phone Interview	Onsite interview
<i>Sophia</i>	3.5	3.6	N/A	4/4	85/100
...					

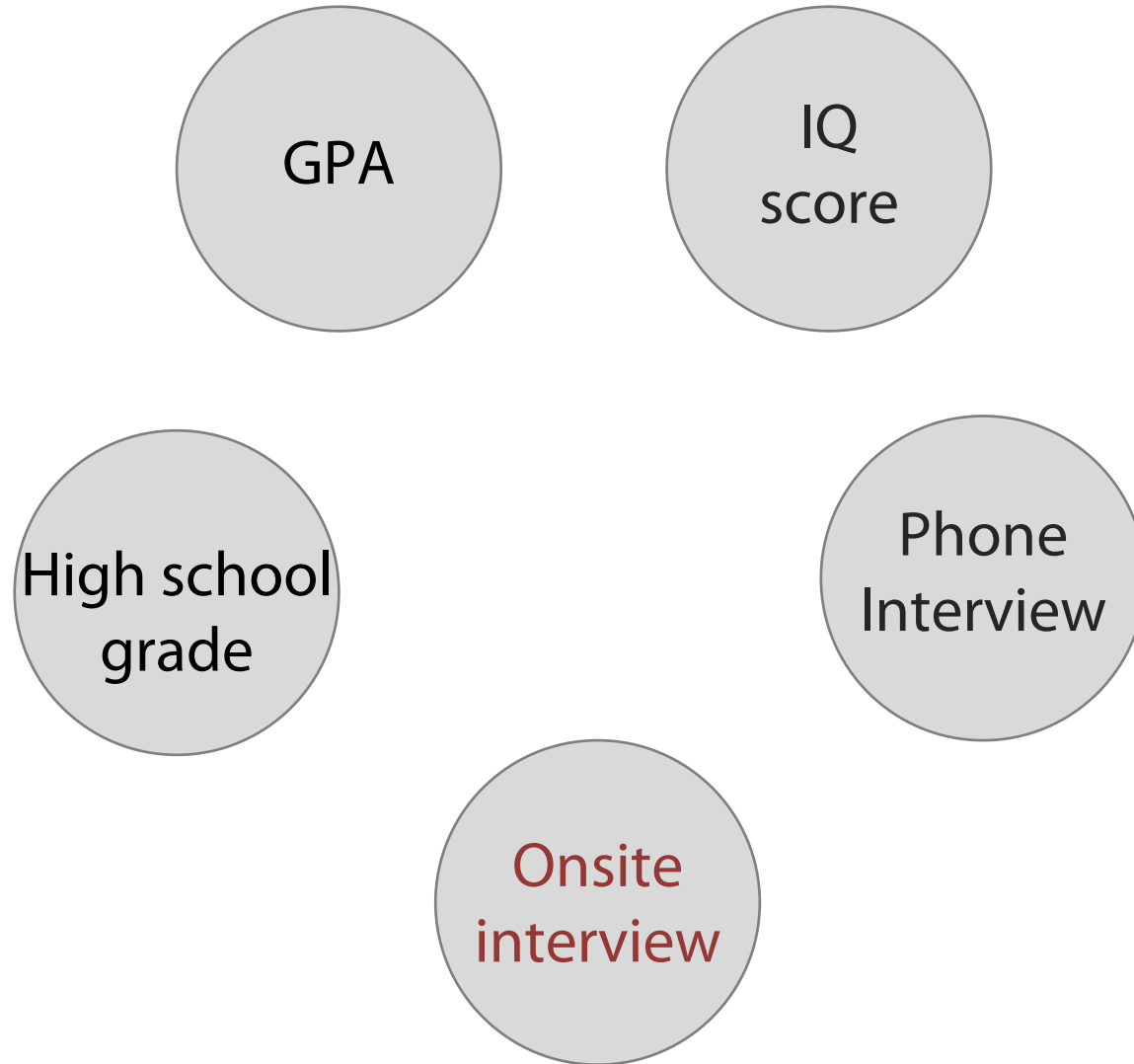
	High school grade	University grade	IQ score	Phone Interview	Onsite interview
<i>John</i>	4.0	4.0	120	3/4	?
<i>Helen</i>	3.7	3.6	N/A	4/4	?
<i>Jack</i>	3.2	N/A	112	2/4	?
<i>Emma</i>	2.9	3.2	N/A	3/4	?

	High school grade	University grade	IQ score	Phone Interview	Onsite interview
<i>Sophia</i>	3.5	3.6	N/A	4/4	85/100
...					

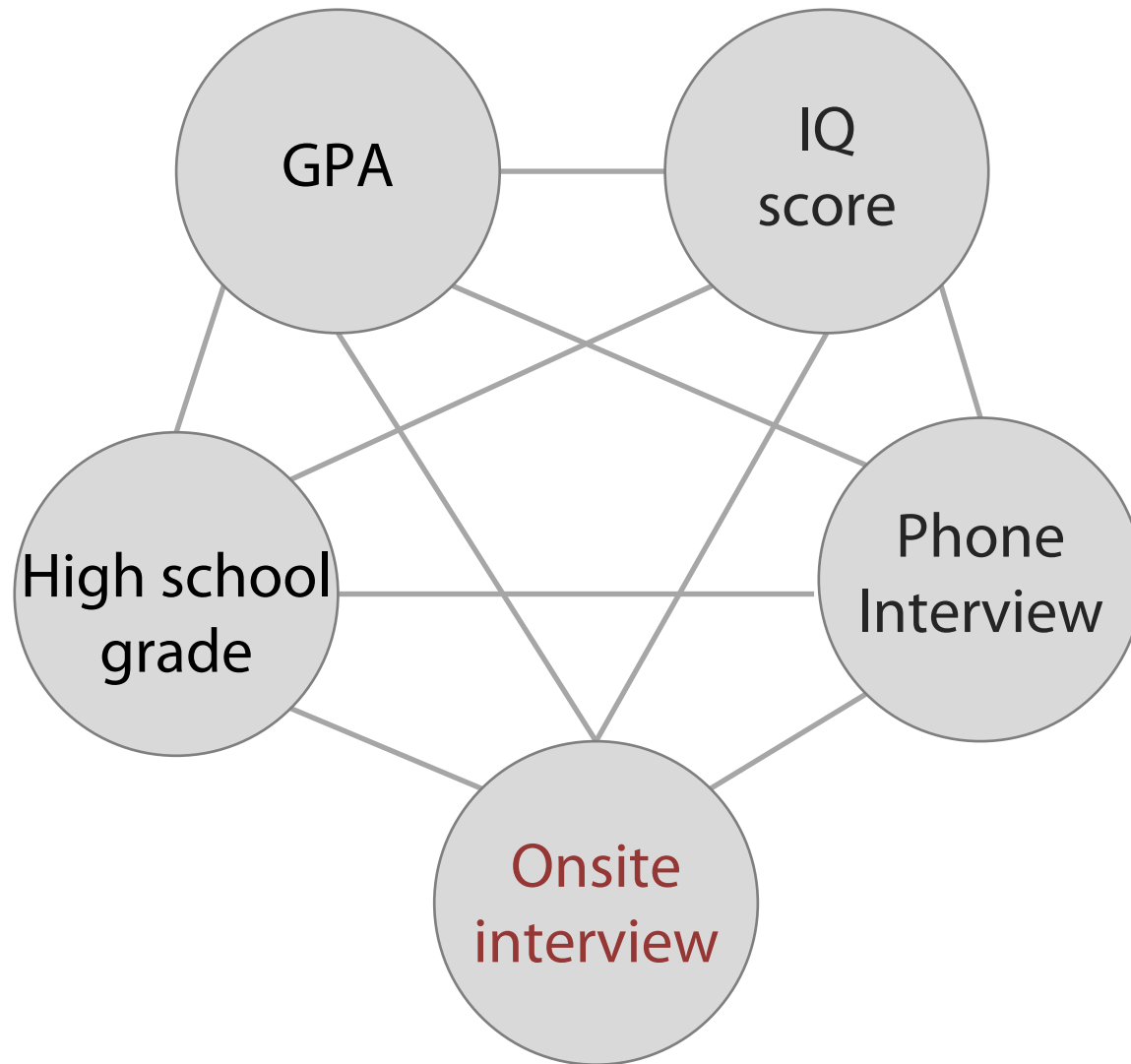
	High school grade	University grade	IQ score	Phone Interview	Onsite interview
<i>John</i>	4.0	4.0	120	3/4	?
<i>Helen</i>	3.7	3.6	N/A	4/4	?
<i>Jack</i>	3.2	N/A	112	2/4	?
<i>Emma</i>	2.9	3.2	N/A	3/4	?



Probabilistic model

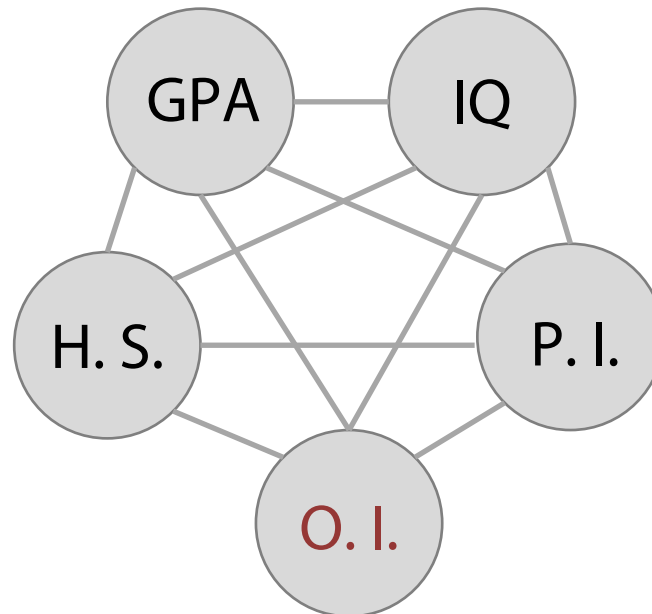


Probabilistic model

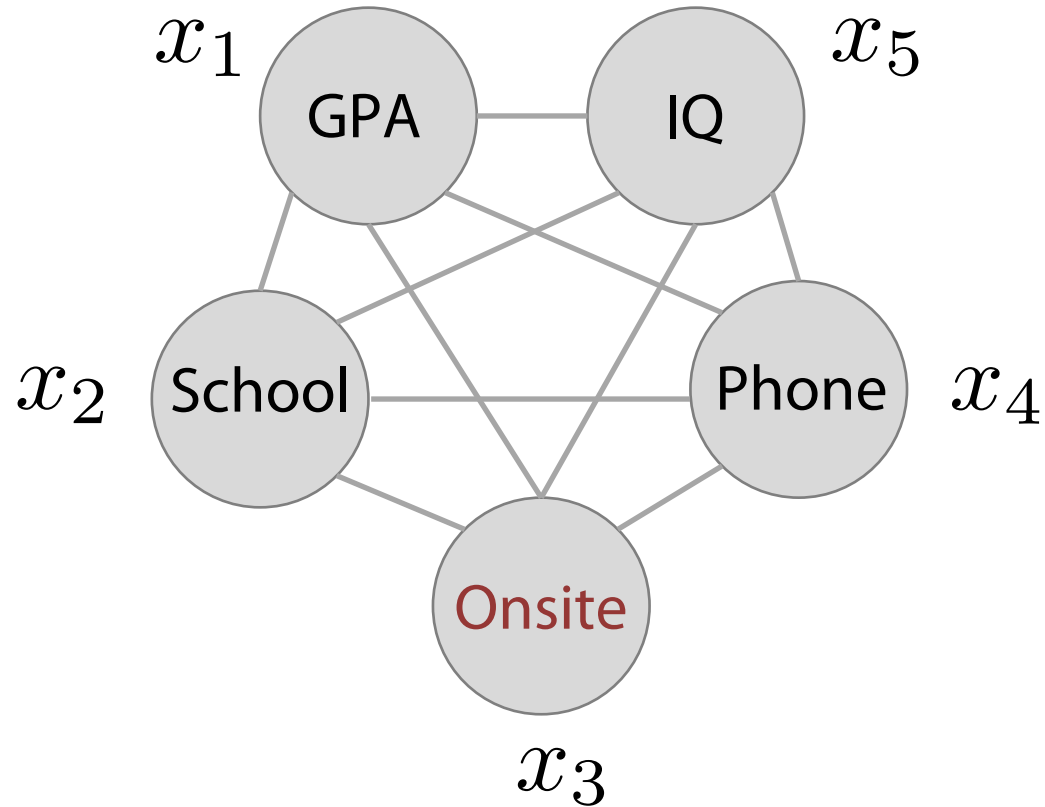


Probabilistic model

High school	GPA	IQ	Phone Interview	Onsite Interview	Probability
1.0	1.0	1	0/4	1/100	0.001
1.0	1.0	1	0/4	2/100	0.0023
...		
4.0	4.0	180	4/4	100	0.000001

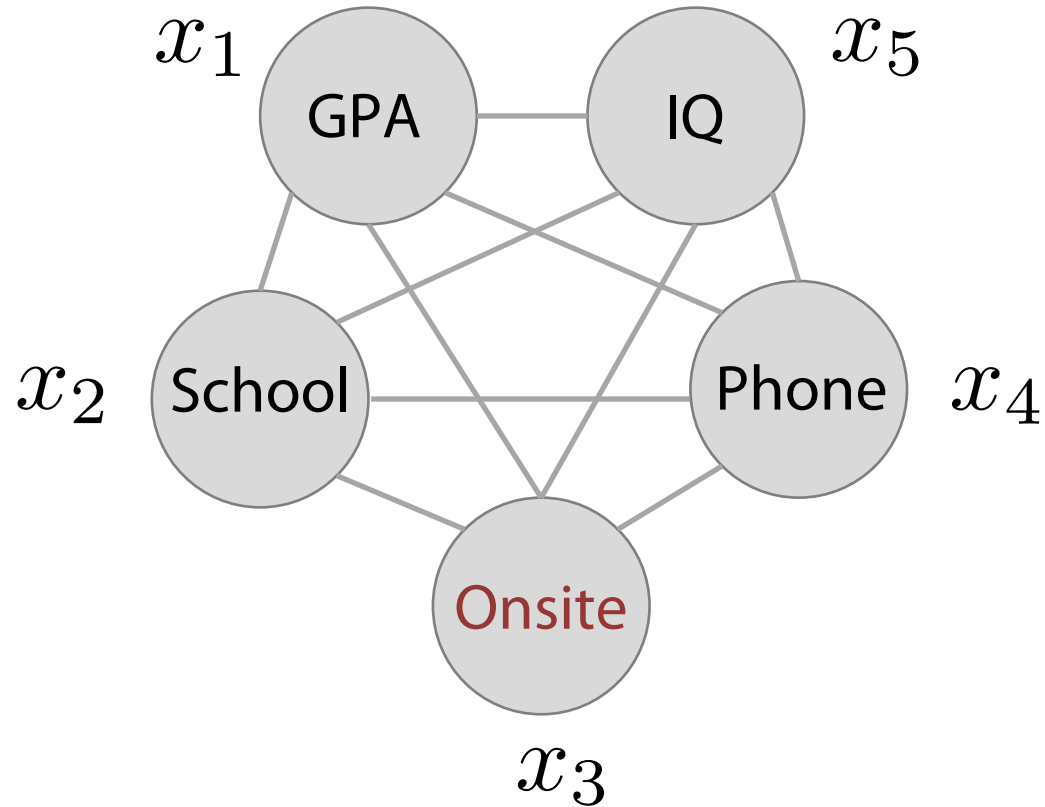


Probabilistic model



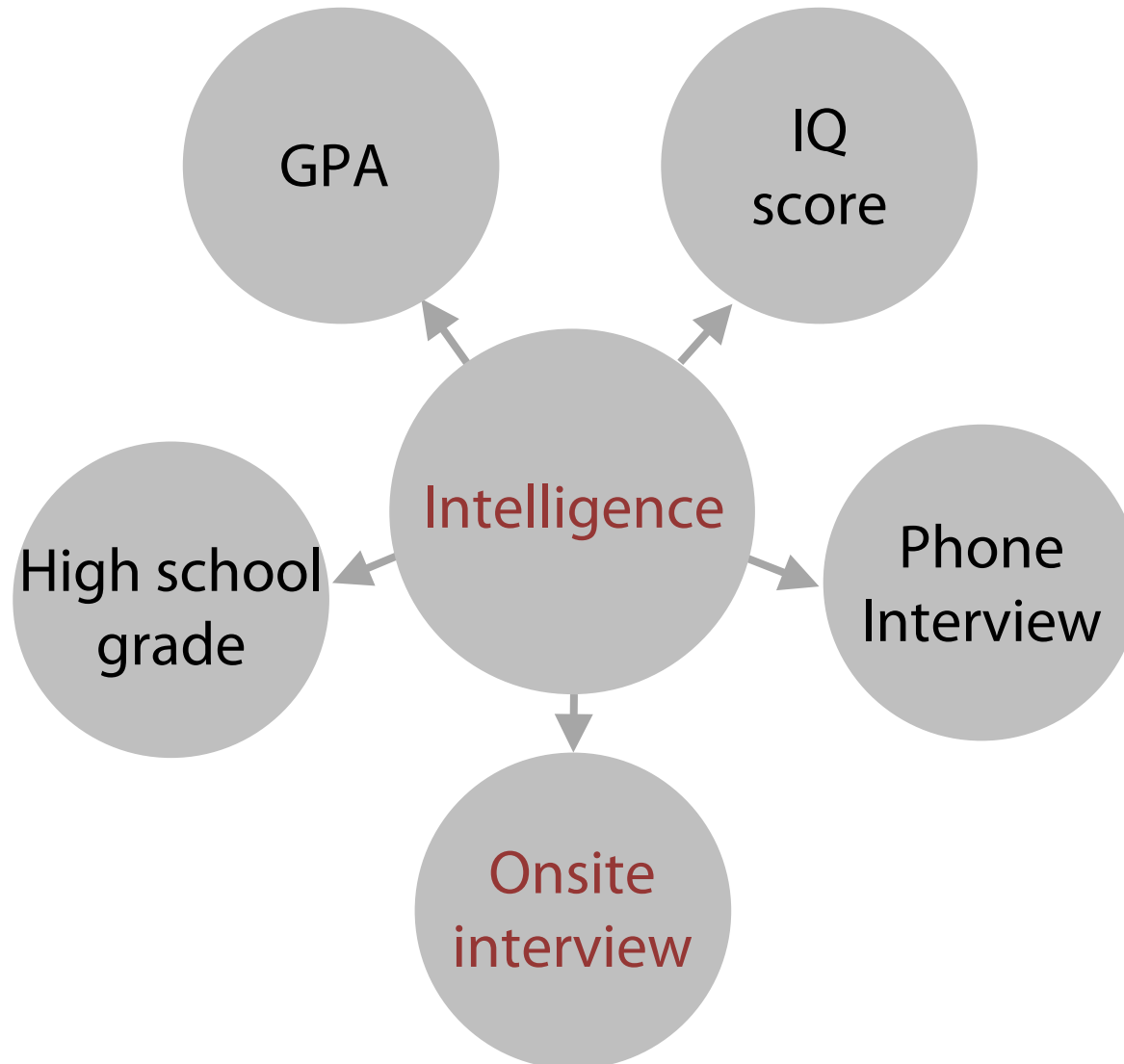
$$p(x_1, x_2, x_3, x_4, x_5) = \frac{\exp(-w^\top x)}{Z}$$

Probabilistic model

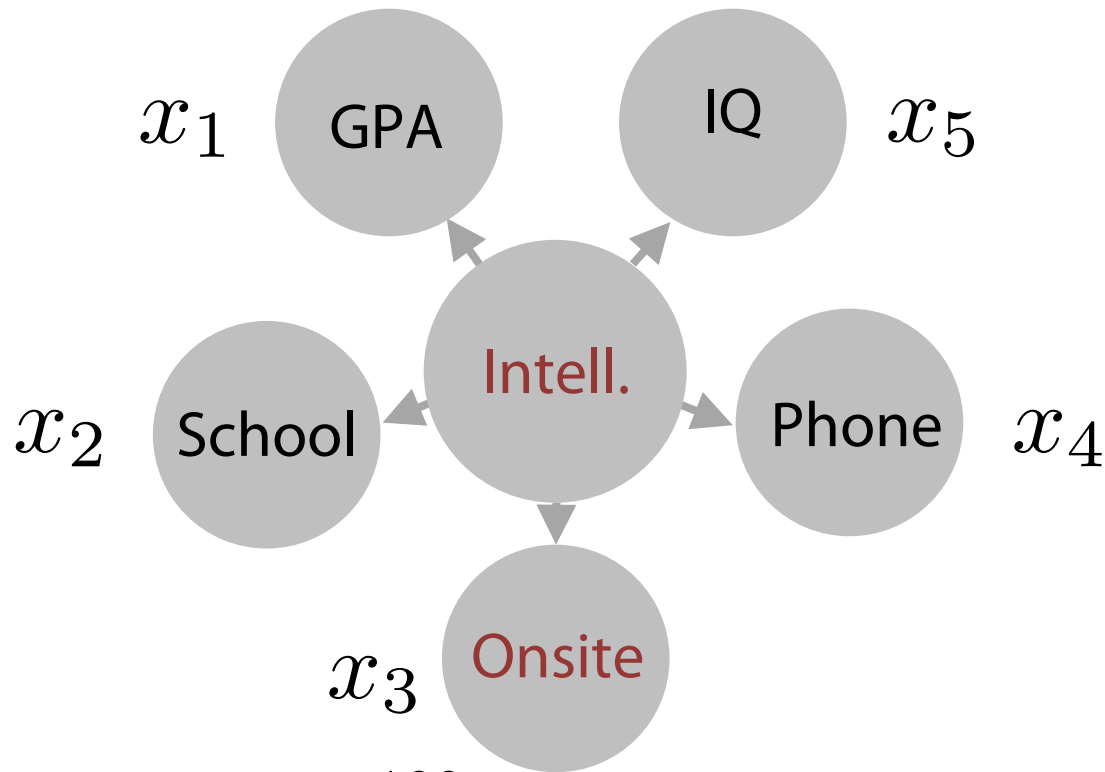


$$p(x_1, x_2, x_3, x_4, x_5) = \frac{\exp(-w^\top x)}{\mathcal{Z}}$$

Probabilistic model

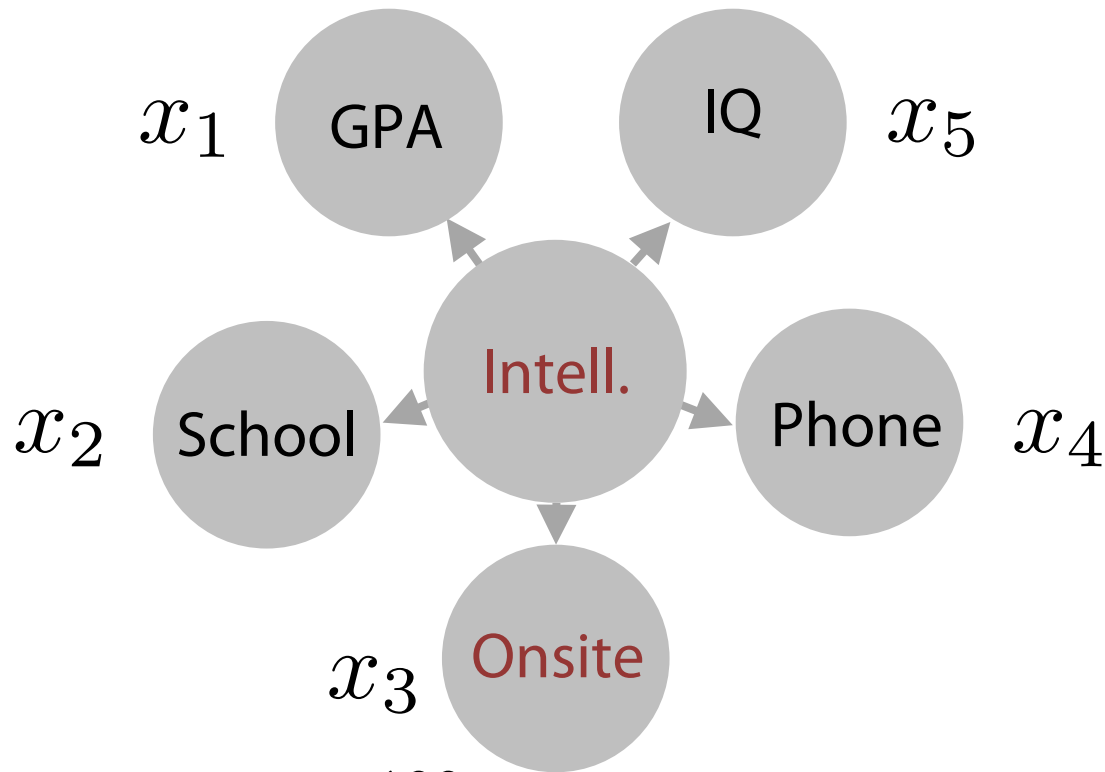


Probabilistic model



$$p(x_1, x_2, x_3, x_4, x_5) = \sum_{I=1}^{100} p(x_1, x_2, x_3, x_4, x_5 \mid I) p(I)$$

Probabilistic model



$$p(x_1, x_2, x_3, x_4, x_5) = \sum_{I=1}^{100} p(x_1, x_2, x_3, x_4, x_5 \mid I) p(I) =$$
$$\sum_{I=1}^{100} p(x_1 \mid I) \dots p(x_5 \mid I) p(I)$$

Latent variable models

Pros:

- Simpler models (less edges)

Latent variable models

Pros:

- Simpler models (less edges)
- Fewer parameters

Latent variable models

Pros:

- Simpler models (less edges)
- Fewer parameters
- Latent variables are sometimes meaningful

Latent variable models

Pros:

- Simpler models (less edges)
- Fewer parameters
- Latent variables are sometimes meaningful

Cons:

- Harder to work with