$$P(x_1, x_2 \dots x_d) = \prod_{j=1}^{d} P(x_j \mid x_{1:j-1})$$

if $\forall i, x_i \in \{0,1\}$ $\Rightarrow$ need 2^d parameters

Simplification

$$P(x_j \mid x_{1:j-1}) = P(x_j \mid x_{pa(j)})$$

$pa(j)$ ↳ parents of j

$x_j \perp x_{-pa(j)} \mid x_{pa(j)}$ → x_j is independent of the other x_i given its parents } conditional independence

$\Rightarrow 2^{(K)+1}$ parameters

(K) ↓ $\#$(Parents of j)

DAG → to represent conditional dependencies

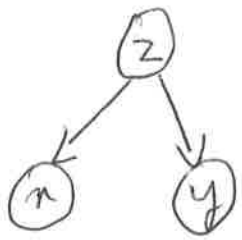
$G = (V, E)$

$V \rightarrow x_i$; $E \rightarrow$ if $(u, v) \in E$ $\Rightarrow u$ is a parent of v

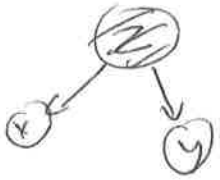
D-separation rules

1) (x) (y) $\Rightarrow x \perp y$

2) (x) → (y) $\Rightarrow x \not\perp y$

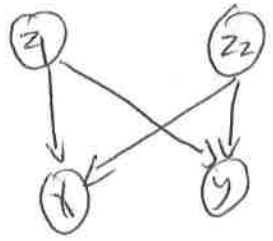
3) 

$\Rightarrow x \not\perp y$

4) 

$\Rightarrow x \perp y \mid z$

5)



$x \perp y \mid z_1, z_2$

$x \not\perp y \mid z_1$

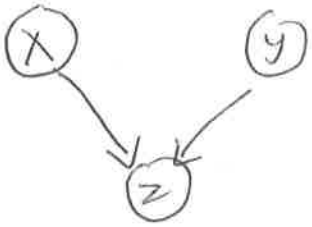
$x \not\perp y \mid z_2$

6)

$x \not\perp y$

$x \perp y \mid z$

7)

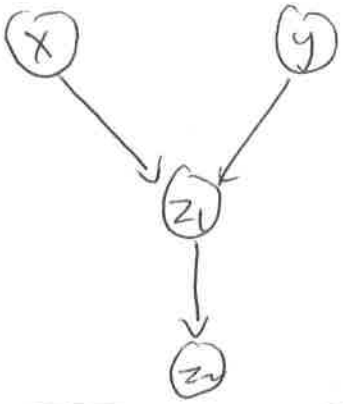


$x \perp y$

$x \not\perp y \mid z$

} V-structure

8)

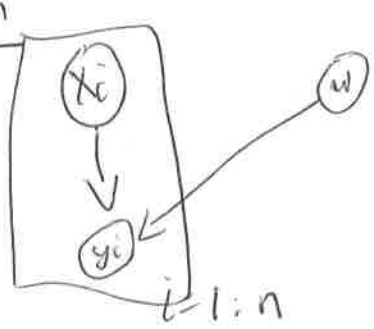


$x \perp y$

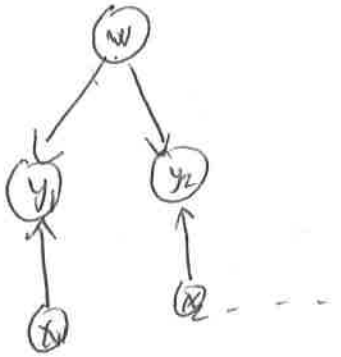
$x \not\perp y \mid z_2$

} V-structure

Plate Notation



$\Rightarrow$



MLE

$y^i \sim N(w^T x^i \mid, 1)$

Project ideas

1) Kaggle / crowdAnalytix
   - US elections 2016 / NCAA basketball / Home depot → search relevance
     SF crime / Yelp photos → multi-label classification
     → multi-class
2) Survey
   - stochastic gradient methods
   - transfer learning
   - Bayesian optimization
   - Multitask learning / multi-label
   - Tensor factorization / spectral methods
   - Sketching techniques
   - classification / anomaly detection for streaming data
3) coding project
   - spectral clustering / ICA / Probabilistic PCA / (code up some theory paper & check how practical it is)
4) Theory
   - matrix completion
   - active learning
   - 2 layer NN's

   ↓
   - submodular maximization
   - bandits
   - online learning
   - randomized SVD
5) check recent AISTATs/NIPS/ICML papers