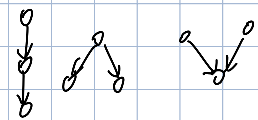
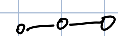
- Basics of probability

  (Axioms, joint, conditional, marginal, independence, Bayes' rule, expectation)

PGM

- Representation
  - BN (DAG)
    - $P(X_1, \dots, X_N) = \prod_{n=1}^{N} P(X_n \mid Par(X_n))$
    - $X \to Y$ if X is a parent of Y
    - Dependencies / Independencies 
      "V-structure", explaining away

    [induction proofs not req]

    ↓ Moralize
  - MN (undirected)
    - $P(X_1, \dots X_N) = \frac{1}{Z} \prod_{k=1}^{K} \phi_k$ ϕ: factor/potential/score
    - MN: $X_i - X_j$ if they cooccur in one factor
    - Factor graph: Bipartite graph of nodes and factors
    - Dependencies / independencies 
  - → CRF: conditional version of MRF $P(Y|X)$

- Inference (exact inference) (undirected)
  - Sum-product, max-product (Commutative semiring not req)
  - Variable elimination: To eliminate a node X
    - Apply evidence potential
    - consider all the factors involving X
    - sum-prod / max-prod
    - result in a factor $\tau(N(X))$
  - Message Passing (Tree of variables, factors, junction tree not req)
    - 2 pass to answer multiple queries about a node
    - $i \xrightarrow{\text{Pass}} j$ if i has received all messages except from j

      $m_{ij} = \sum_{\substack{\text{variables} \\ \text{to eliminate}}} \phi_{local} \prod_{k \in N(i) \setminus j} m_{k \to i}$

- Learning (Fully observable BN)
  - MLE decomposes to each node conditioned on the values of its parents.
  - Categorical distribution
    - MLE: $\hat{\pi}^{(MLE)} = \underset{\pi}{\operatorname{argmax}} P(\mathcal{D}|\pi)$ $\hat{\pi}_k^{(MLE)} = \frac{N_k}{N}$
    - MAP: $\hat{\pi}^{(MAP)} = \underset{\pi}{\operatorname{argmax}}\ p(\pi) p(\mathcal{D}|\pi)$ add-$(\alpha - 1)$ smoothing

      $p(\pi) = Dir(\pi; \alpha)$
    - Bayesian: $p(x|\mathcal{D}) = \int p(x|\pi) \underbrace{p(\pi|\mathcal{D})}_{\propto\, p(\pi) \cdot p(\mathcal{D}|\pi)} d\pi$ add-α smoothing
    - learning α: MLE, MAP, Bayesian (hierarchical Bayes)

(Memorizing formulas of $\mathcal{N}$, Dir is not req.)

(Computing $\int$ is not req unless taught in lec)

Emphasis: small derivations on standard knowledge. Unimportant: { memorization; tricky derivation steps }