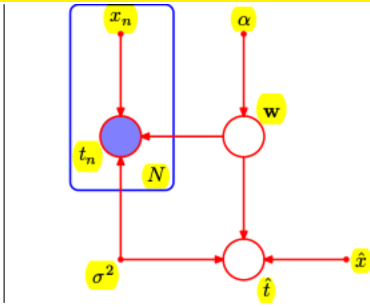
1. **Directed graphs** are useful for expressing causal relationships between random variables, whereas **undirected graphs** are better suited to expressing soft constraints between random variables. For the purposes of solving inference problems, it is often convenient to convert both directed and undirected graphs into a different representation called a factor graph.

2. We consider DAGs, and general factorization is $p(\boldsymbol{x}) = \prod_{i \in n} p(x_i | \text{pa}(x_i))$.

3. Suppose we are given a new input value $\hat{x}$ and we wish to find the corresponding probability distribution for $\hat{t}$ conditioned on the observed data. We get the figure (it has the *plate* notation for N variables.) and the equation

$$p(\hat{t}, \boldsymbol{t}, w | \hat{x}, \boldsymbol{x}, \alpha, \sigma^2) = \prod_{i \in [n]} p(t_1 | x_i, w, \sigma^2) p(w|\alpha) p(\hat{t} | \hat{x}, w, \sigma^2).$$



4. **Ancestral Sampling**: Topological sort the DAG, and start from parents and generate values for children. To draw a sample from the distribution $p(x_2, x_4)$, sample from full joint distribution, and retain values x_2, x_4 and discard the rest.

5. To **reduce number of parameters**, reduce the number of links in graph, by either sharing parameters or removing the links. In case of discrete random variables, it can range from fully connected DAG ($K^M - 1$ params.) or a fully disconnected DAG ($M(K-1)$ params.). Similar things can be done for Gaussian random variables .

6. **Conditional Independence**: a is conditionally independent of b given c if $p(a|b,c) = p(a|c)$ or $p(a,b|c) = p(a|c)p(a|c)$.

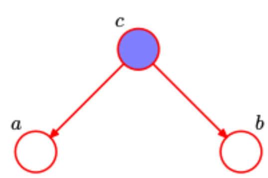


Fig1: Node c is said to be *tail-tail* wrt path from a to b.

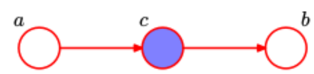


Fig2: Node c is said to be *head-tail* wrt path from a to b.

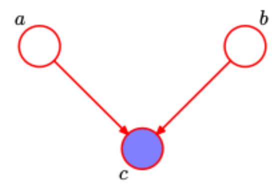


Fig3: Node c is said to be *head-head* wrt path from a to b.

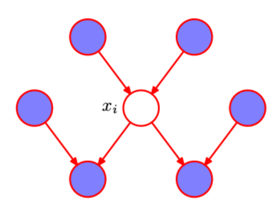
7. **Three examples**: The first two figures have conditional independence between a and b, and third figure has marginal independence between a and b, when c is unobserved (shaded c means observed c). In Fig3, c explains away a and b. A tail-tail/head-tail node leaves a path unblocked unless it is observed in which case it blocks the path. A head-head node blocks a path if it is unobserved, but if the node, and/or at least one of its descendants, is observed the path becomes unblocked.

8. *d*-**Separation**: Consider a directed graph in which A, B, C are nonintersecting subsets of nodes. Consider *all* possible paths from any node in A to any node in B. Any such path is said to be *blocked* if it includes a node such that either
   (a) arrows on the path meet either head-tail/tail-tail at a node, and the node is in the set C, or
   (b) arrows meet head-head at a node, and neither the node, nor any of its descendants, is in the set C.

   If all paths are blocked, then A is said to be d-**separated** from B by C, and the joint distribution over all of the variables in the graph will satisfy $A \perp\!\!\!\perp B | C$.

9. **Naive Bayes Model**: assume conditional independence assumptions to simplify the model structure. Useful assumption for high-dimn. data, to lessen the parameters.

10. **Markov Blanket**: of node $\boldsymbol{x}_i$ is the minimal set of nodes that isolates $\boldsymbol{x}_i$ from the rest of the graph. $p(\boldsymbol{x}_i | \boldsymbol{x}_{j \neq i})$ will have factors that will be the conditional distribution $p(\boldsymbol{x}_i | \text{pa}(\boldsymbol{x}_i))$ and the conditional distributions for nodes $\boldsymbol{x}_k$ such that node $\boldsymbol{x}_i$ is in the conditioning set of $p(\boldsymbol{x}_k | \text{pa}(\boldsymbol{x}_k))$. Note that there are co-parents of children of $\boldsymbol{x}_i$ included due to *head-head* relationship in the figure.



Markov blanket of $\boldsymbol{x}_i$