GRAPHICAL MODEL

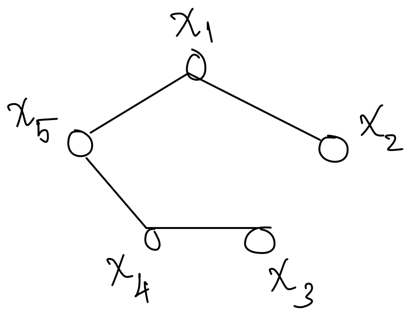
- Prequesites: Multivariate Normal, Precision Matrix

Main idea:

- Recall that Precision Matrix measure <u>conditional independence</u>
- We can actually represent the Precision Matrix using a graph.

  Given this sparse precision matrix:

$$\begin{array}{c} \\ x_1 \\ \vdots \\ \\ x_5 \end{array} \begin{array}{c} x_1 \quad \cdots \quad x_5 \\ \begin{bmatrix} 2 & 1 & & & 1 \\ 1 & 2 & & & 1 \\ & & 2 & 1 & \\ & & 1 & 2 & 1 \\ 1 & 1 & & 1 & 2 \end{bmatrix} \end{array} \Longleftrightarrow$$



We can read Markov (conditional independence) from the graph, like:

i) $x_1 \perp (x_3, x_4) \mid (x_2, x_5)$

Markov blanket

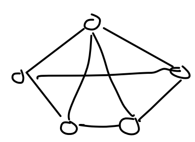
Condition on (x_2, x_5) will make x_1 independent (x_3, x_4)

ii) i, j disconnected $\Rightarrow x_i \perp x_j \mid x_{\neg ij}$

If 2 nodes are disconnected, then they are independent conditioned on any path between them (example: x_2 and x_3)

Why not use Covariance Matrix:

- Recall that the Covariance Matrix measure the marginal independence between variables (features), it is pairwise
- If we were to graph out the Covariance Matrix, we would get a fully connected graph (if all variables are correlated)



This is not very meaningful.

How to find (estimate) the graph?

- Use a technique called <u>Graphical Lasso</u>. The main idea is to use maximum likelihood estimation with some penalty to encourage sparsity
- Mathematically:

$$\hat{Q} = \underset{Q}{\operatorname{argmax}} \log P(D \mid Q) - \lambda \phi(Q)$$

where $D = \{x_i\}_{i=1}^{n} \sim N(0, \Sigma)$, $Q = \Sigma^{-1}$

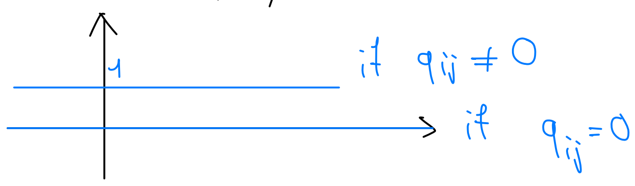
- On the penalty:

  It "strict" definition is the number of edges, called L_0 norm

  $$\phi(Q) = \|Q\|_0 = \sum_{i \neq j} I(q_{ij} \neq 0) \quad = \# \text{ of edges}$$

  $$= \sum_{i \neq j} q_{ij}^0 = \begin{cases} 0 & \text{if } q_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

  $\Rightarrow$ This function is not convex, so its hard to maximize



  So we relax the definition and use L1 norm instead:

  $$\phi(Q) = \|Q\|_1 = \sum_{i \neq j} |q_{ij}|$$

  $$= \begin{cases} 0 & \text{if } q_{ij} = 0 \\ |q_{ij}| & \text{otherwise} \end{cases}$$

  $\Rightarrow$ This function is convex, and can be maximized