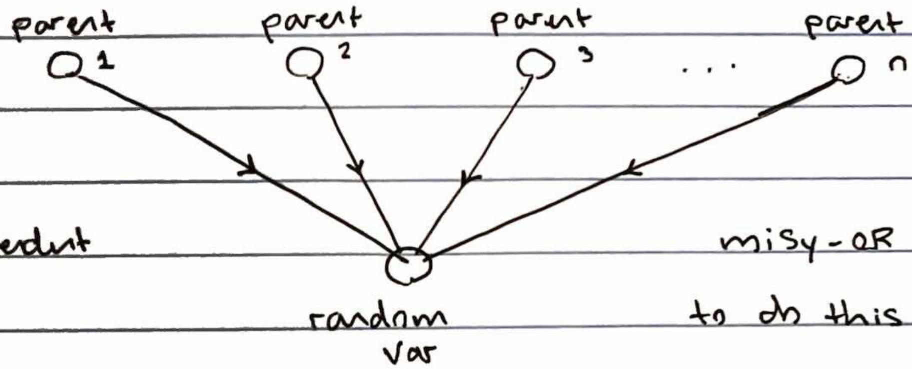
through the process of combining these variables we greatly reduce the data required for training? ~ this seems wrong



where each parent is independent

noisy-OR function to do this.

"ICI can be very restrictive and easily violated"
this may not work in real-life!

Qualitive Influences, how a change in one variable effects a change in another. Example is "monotonicity" x affects y in the same way data is monotonic x up y up. Combining QI with context specific independencies is a special case of "isotonic regression"

The point of the paper:

- Current QI can handle monotonicities we extend this to allow for synergistic interactions.
- Combine Synergistic and monotonic interactions with the concept of ICI i.e. they treat set of monotonicities and set of synergies as independent and combine these with noisy-OR

Synergy constrained Noisy-OR leads to more accurate models in the presence of smaller datasets.

parameters of Bayes Network θ_{ijk} where $\{i \in 1,2,\dots n,\ j \in 1,2,\dots v_i,\ k \in 1,2,\dots r_i\}$

the conditional probability of X_i is the kth vale given the jth configuration of it's parents $X_i = P(X_i[k] \mid x_i'[j])$ in the paper they write this as $P(X_i^k \mid pa_i^j)$ r_i is the number of states of X_i, Pa_i is the parent set of X_i the number of configurations of Pa_i is $v_i = \prod_{X_t \in Pa_i} r_t$ j is the index of pa_i's configuration.

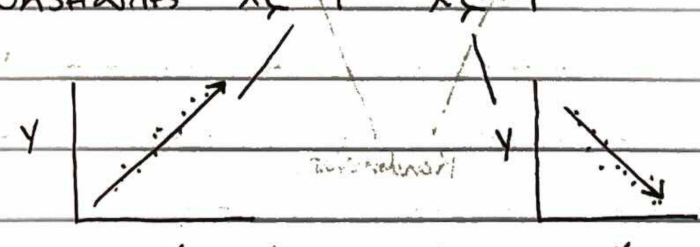
General Form of Monotonic Constraint

$P(X_i \le k_c \mid X_c^m, C_i^n) \ge P(X_i \le k_c \mid X_c^{m+1}, C_i^n)$

where $k_c \in \{1,2,\dots,r_i-1\}$, $m \in \{1,2,\dots,r_c-1\}$

$X_c^m \le X_c^{m+1}$ C_i^n repersents all possible configurations of X_i parent other than X_c n is the index

notation for monotonic constraints $X_{\overrightarrow{M+}}Y$ $X_{\overrightarrow{M-}}Y$



Assume we have a monotonic constraint

$P(X_i \le K_c \mid pa_i^{j_2}) \le P(X_i \le K_c \mid pa_i^{j_1})$

the constraint function δ with margin ϵ

$$\delta = P(X_i \le K_c \mid pa_i^{j_2}) - P(X_i \le K_c \mid pa_i^{j_1}) + \epsilon$$

the penalty function is $P_{j_1 j_2}^{i,k_c} = I_{(\delta > 0)}\delta^2$ ($I = 1$ when $\delta > 0$ else $I = 0$)

2.2 Noisy-OR

if there are n independent causes $\{X_1 \ldots X_n\}$ for Y assuming Y is binary for simplicity then the distrobution y=1, $P(y=1 \mid X_1 \ldots X_n)$ is given by: $1 - \prod_i P(y=0 \mid X_i = x_i) = P(y=1 \mid X_1 \ldots X_n)$

Y will take a value of 1 unless there's an inhibition the probability of this is $P(y=0 \mid X_i = x_i)$ these effects are assumed independent $1-q_i$ for the ith parent. we can re-write Noisy-OR like so:

$$1 - \prod_i (1-q_i)^{x_i}$$

3. Qualitative Constraints Synergies

synergy is feature combination when both features have the same monotonic effect. i.e. given X_1 and X_2 ~~dependently~~ influence Y. $X_1, X_2 \xrightarrow{S+} Y$ X_1 inc. inc. Y X_2 inc. inc. Y the example given is blood pressure goes up and chelestorol goes up this increases risk for heart attack synergisticly. Parents cannot have different monotonic relationships with Y.