# Weak Supervision

INDEP CASE → A simple "trick" for latent variables

Correlations → Inverse Covariance ⟷ Graphs (Gaussians)

Graphical models

**Nuggets:** Method of Moments

USED IN Crowd Sourcing, WS, Classical stats (EM)

GIVEN: $x^{(1)} \dots x^{(n)} \in \mathbb{R}^d$

$\lambda_i : \mathbb{R}^d \to \{-1, 1\} \cup$ "ABSTAIN" for $i = 1 \dots m$

find $P(y^{(i)} \mid \bar{\lambda}, x^{(i)})$

Given

| DATA | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $y$ | |
|---|---|---|---|---|---|
| | OBSERVE | | | Ground truth (UNOBSERVED) | |
| $x^{(1)}$ | 1 | 1 | 1 | 1 | $y^{(1)}$ |
| $x^{(2)}$ | 1 | -1 | 1 | -1 | $y^{(2)}$ |
| $\vdots$ | | | | | |
| $x^{(n)}$ | -1 | -1 | -1 | -1 | $y^{(n)}$ |

IDEA: $\lambda_i$ Are noisy voters / function (INACCURATE)

MODEL 0: No Abstains, $\lambda_i(x) \in \{-1, 1\}$, INDEPENDENT ERRORS

"EACH LABELER has A hidden accuracy, MAKES AN ERROR w/ 1-p prob"

for EACH Example $x^{(i)}$ AND $\lambda_j$

$\lambda_j(x^{(i)}) = y^{(i)}$ w/ PROB $p_j$ "$\lambda_j$ IS Right"

$\lambda_j(x^{(i)}) = -y^{(i)}$ w/ Prob $1-p_j$ " IS wrong"

WE NEED TO ESTIMATE $p_j$

$$P(\lambda_j(x) = 1 \mid y = 1) = P(\lambda_j(x) = -1 \mid y = -1)$$

Goal Compute $p(y^{(i)} \mid \bar{\lambda}, x)$

for $m \geq 3$

DEF: $\mathbb{E}[\lambda_j y] = \underbrace{p_j \cdot 1}_{\text{Agree}} + \underbrace{(1-p_j)(-1)}_{\text{disagree}} = 2p_j - 1$

define $a_j = 2p_j - 1$

$$\mathbb{E}[\lambda_i \lambda_j] = \begin{cases} 1 & \text{if } i = j \quad \lambda_j \in \{-1, 1\} \\ & \text{o.w.} \\ 1\,(p_i p_j + (1-p_i)(1-p_j)) & \text{Agree} \\ + -1\,((1-p_i)p_j + p_i(1-p_j)) & \text{disagree} \end{cases}$$

$= a_i \cdot a_j$

form A MATRIX $M \in \mathbb{R}^{m \times m}$ $\quad M_{ij} = \mathbb{E}[\lambda_i \lambda_j]$

NB: WE CAN ESTIMATE from DATA (unlike $y$)

## CRAZY Algorithm

OBSERVE $\quad M_{ij} M_{jk} = a_i a_j^2 a_k$

$$\frac{M_{ij} M_{jk}}{M_{ik}} = a_j^2 \quad \text{"compute triples"}$$

WE HAVE SOLVED UPTO SIGN BASED ON OBSERVED $M_{ij}$

$$\text{Sign}(M_{ij}) = \text{sign}(a_i) \cdot \text{sign}(a_j)$$

ONCE WE KNOW EVEN A SINGLE READER'S SIGN

<u>However</u> $a, -a$ Are solutions

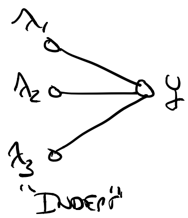
<u>WEAKER Solution</u> $\sum_i a_i > 0$, BREAKS Symmetry

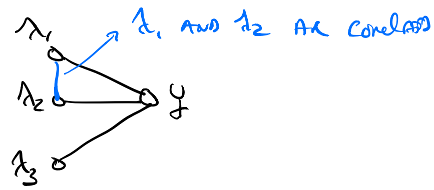
THIS IS A SIMPLE Solution to AN em like Problem. ☐

## WHAT IF $\lambda_i$ ARE CORRELATED?

In general, hopeless.

A graphical models




$$\mathbb{E}[\lambda_2 \lambda_3 \mid y] = \mathbb{E}[\lambda_2 \mid y]\, \mathbb{E}[\lambda_3 \mid y]$$

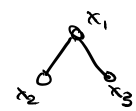
$$G = (\underbrace{V}_{\text{NODES}}, \overbrace{E}^{\text{EDGES}}) \qquad (1, y) \notin E \qquad V = \{\lambda_1, \dots, \lambda_m\}$$

## Nugget Covariance Structure

$$X_1 \sim N(0,1)$$

$$X_2 = X_1 + \epsilon_2 \qquad \epsilon_2 \sim N(0,1)$$

$$X_3 = X_1 + \epsilon_3 \qquad \epsilon_3 \sim N(0,1)$$




**MEAN**

$$\mathbb{E}[X_1] = 0 \qquad \mathbb{E}[X_2] = \underbrace{\cancel{\mathbb{E}[X_1]}}_{\to 0} + \underbrace{\cancel{\mathbb{E}[\epsilon_2]}}_{\to 0} = 0 = \mathbb{E}[X_3]$$

**COVARIANCE**

$$\mathbb{E}[X_1^2] = 1 \qquad \mathbb{E}[X_2^2] = \mathbb{E}[(X_1+\epsilon_2)^2] = \overbrace{\mathbb{E}[X_1^2]}^{1} + \overbrace{\mathbb{E}[\epsilon_2^2]}^{1} + \underbrace{\cancel{2\,\mathbb{E}[X_1\epsilon_2]}}_{\to 0} = 2$$

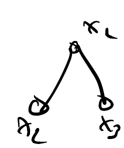
$$\mathbb{E}[X_1X_2] = \overbrace{\mathbb{E}[X_1^2]}^{1} + \cancel{\mathbb{E}[X_1\epsilon_2]} = 1 = \mathbb{E}[X_1X_3]$$

$$\mathbb{E}[X_2X_3] = \mathbb{E}[(X_1+\epsilon_2)(X_1+\epsilon_3)]$$

$$= \mathbb{E}[X_1^2] + \cancel{\mathbb{E}[X_1\epsilon_2] + \overset{0}{\mathbb{E}[X_1\epsilon_3]} + \mathbb{E}[\epsilon_2\epsilon_3]}$$

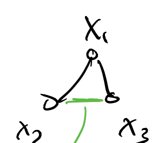
$$= 1$$

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$




NO OBVIOUS!

$$\Sigma^{-1} = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$




0 when no edge is present!

The fact is AWESOME, let's prove Gaussians!

(Wainwright & Loh 2014 about discrete case

Ratner 2018)

A probability distribution $p: \mathbb{R}^d \to [0,1]$ factorizes

or agrees w/ a graph $G = (V, E)$ if

$$p(x) = C_0 \cdot \prod_{e=(i,j)\in E} p_e(x_i, x_j) \cdot \prod_{i \in V} p_i(x_i)$$

$C_0 \to$ NORM CONSTANT

OTY $\quad p(x) = C \cdot \exp\{-x^T \Sigma^{-1} x\}$ $\quad$ Gaussians?

(*)

define $A = \Sigma^{-1}$ (notation)

$$\log p(x) = \log C - \underbrace{x^T A x}_{\sum_{i,j} A_{ij} x_i x_j}$$

If $p$ factors wrt $G = (V, E)$

$$\underset{(**)}{\log p(x)} = \log C_0 + \sum_{e=(i,j)\in E} \log p_e(x_i, x_j) + \sum_{i \in V} \log p_i(x_i)$$

Consider $(i,j) \notin E$ $\quad \frac{\partial}{\partial x_i \partial x_j} \log p(x)$

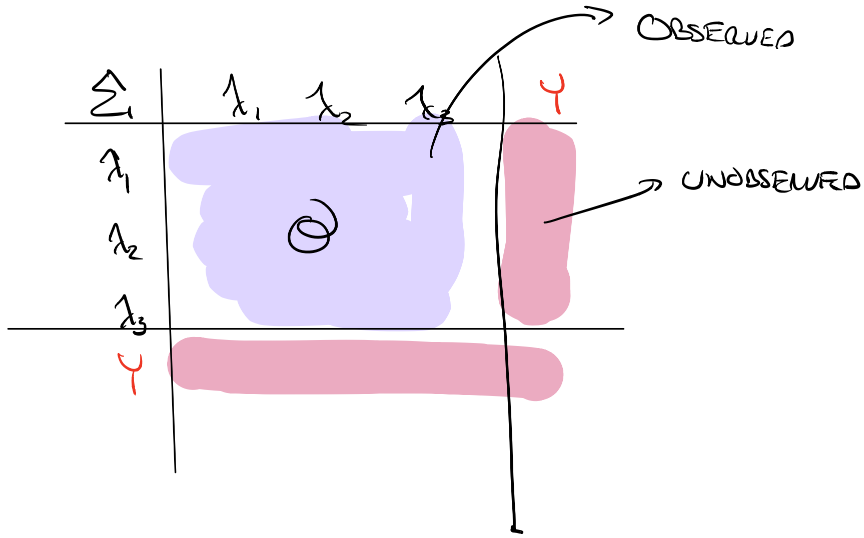
$$\frac{\partial}{\partial x_i \partial x_j} (*) = A_{ij} + A_{ji} = 2A_{ij} \quad \text{(covariance symmetric)}$$

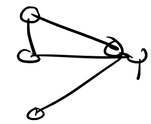
$$\frac{\partial}{\partial x_i \partial x_j} (**) = 0 \qquad = 2A_{ij}$$

$\therefore \quad \Sigma^{-1}_{ij} = 0$ whenever $(i,j) \notin E$.

IT'S NOT quite as clean for discrete r.v.

## Back to our Problem



Assume we know graphical structure $G=(V,E)$

then from above $(i,j) \notin E \Rightarrow (\Sigma^{-1}_{ij}) = 0$

IDEA: if there aren't too many edges, should be able

Accuracy and Correlation strength

Know $O$ entries — Do not $O$ entries? But I can measure

① $(\hat\Sigma^{-1})_O = (O - uu^T)^{-1}$ (Block inversion lemma)

$O = \{1,2,3\}$

② $(O - uu^T)^{-1} = O^{-1} + zz^T$ $\quad z = \frac{O^{-1}u}{\sqrt{1-u^TOu}}$

solve for

$$\underbrace{(\Sigma^{-1})_{ij}}_{(i,j)\notin E} = (O^{-1})_{ij} + z_i z_j$$

$= 0 \Rightarrow -O^{-1}_{ij} = z_i z_j$ linear system

$\log (O^{-1}_{ij})^2 = \log z_i^2 + \log z_j^2$

$\Rightarrow$ linear system in logs

∴ WE CAN RECOVER whenever this full RANK!

+ This technique "Adding variables" (WHAT WE DID w/ 4) IS Powerful ⟹ Higher RANK Solves

+ LEARN Graph Structure "guess" (w/ SDP)

## RECAP

+ "METHOD of moments"
+ Nugget Graphs ⟺ Probability (graphical models)
+ Application WEAK Supervise