Weak Supervision Nuggets

- Indep case → Simple Estimation task
- Correlations → Inverse Covariance & Graph

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

$\lambda_1 \dots \lambda_m \; : \; \lambda_i : \mathbb{R}^d \to \{-1, 1\} \cup \{\text{Abstain}\}$

Find $P(y^{(i)} \mid \bar{\lambda}, x^{(i)}) \quad y \in \{-1, 1\}$

Idea: λ_i is a noisy voter/function (inaccurate)

λ_1 : "the classifier says yes"

λ_3 : "name in DS"

... Programmatic labels ..

Model 0: No abstains, Independent errors.

Each labeler λ_i has hidden accuracy. Fix x & y → unobserved label

w/ Prob P_i $\quad \lambda_i(x) = y \quad$ "λ_i is right"

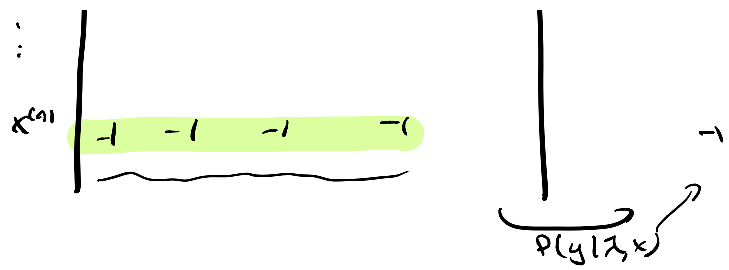
$1 - P_i$ $\quad \lambda_i(x) = -y \quad$ "λ_i is wrong"

Sadly, we don't see y.

$$P(\lambda_i(x) = 1 \mid y = 1) = P(\lambda_i(x) = -1 \mid y = -1) = P_i$$

Given

Data	λ_1	λ_2	λ_3	.. λ_m	Unobserved y
$x^{(1)}$	1	1	1	1	1
$x^{(2)}$	1	-1	1	1	-1



① $\mathbb{E}[\underset{\in\{-1,1\}}{\lambda_i} Y] = P_i \cdot 1 + (1-P_i)\cdot -1 = 2P_i - 1$ def a_i

$a_i \triangleq 2P_i - 1$

$a_i \in [-1, 1]$

② $\mathbb{E}[\lambda_i \lambda_j] = 1$ if $i = j$ $(\mathbb{E}[\lambda_i^2] = \mathbb{E}[1])$

③ $\mathbb{E}[\lambda_i \lambda_j]$ $i \neq j$

$= P_i P_j \cdot 1 + (1-P_i)(1-P_j)\cdot 1$ is true

$P_i(1-P_j)(-1) + (1-P_i)P_j$ -1

$= a_i \cdot a_j$

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

NB: we can estimate M from data (observed) unlike y.

"Agreements & disagreements" no need for y.

Simple Algorithm

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

$\frac{M_{ij} M_{jk}}{M_{ik}} = a_j^2$ solve "upto the sign of a_i"

we know magnitude, not sign

M_{ij} is observed $M_{ij} = a_i a_j$ if I know sign(a_i)

sign(M_{ij}) = sign(a_i) sign(a_j) $\Rightarrow$ sign(a_j)

$\Rightarrow$ $a, -a$ both are solutions.

$\sum_i a_i > 0$, breaks symmetry "labelers are not all worse than random or adversarial"

· What if $M_{ij} = 0 \Rightarrow a_i = 0$ or $a_j = 0$

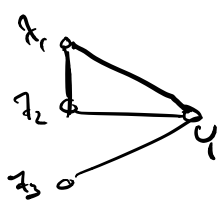
$\Rightarrow a_i = 0 \Rightarrow 2p_i - 1 = 0 \Rightarrow p_i = \frac{1}{2}$ $\quad$ p_i are bounded away from $\frac{1}{2}$

$|p_i - \frac{1}{2}|$ distance

Recap: Simple solution to "EM-like"

Symmetry $\Rightarrow$ had to be in model

What if the labelers are correlated



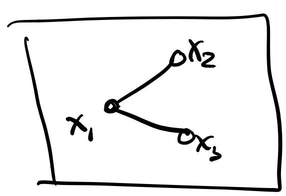
$$E[\lambda_1 \lambda_2 | Y] = E[\lambda_1 | Y] E[\lambda_2 | Y]$$

"if there edge between two r.v. (nodes)"

$\Rightarrow$ then they are not independent

$(i,j) \notin E$ holds for any i,j.

Nugget Structure of Inverse Covariance Matrices



$X_1 \sim N(0,1)$

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

$X_2 \sim N(X_1, 1) \quad \equiv \quad X_3 = X_1 + \epsilon_3 \qquad \epsilon_3 \sim N(0,1)$

$X_3 \sim N(X_1, 1)$

In $E[X_i] = 0 \qquad E[X_2] = E[X_1] + E[\epsilon_2] = 0$

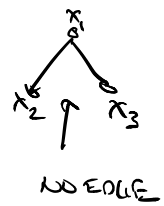
$E[X_3] = E[X_1] + E[\epsilon_3] = 0$

2. $\mathbb{E}[x_1^2] = 1$ $\quad \mathbb{E}[x_2^2] = \mathbb{E}[(x_1+\epsilon_2)^2] = \mathbb{E}[x_1^2] + \mathbb{E}[\epsilon_2^2] + \cancel{0}$

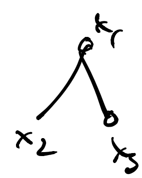
$\mathbb{E}[x_3^2] \qquad = 2$

3. $\mathbb{E}[x_1 x_2] = \mathbb{E}[x_1^2] + \cancel{\mathbb{E}[x_1\epsilon_2]}^{\,0} = 1$

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



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



$V = \{x_1, x_2, x_3\}$

$E = \{(x_1, x_2), (x_1, x_3)\}$

We say a probability distribution $p: \mathbb{R}^d \to [0,1]$ <u>factorizes</u> or <u>agrees</u> with a graph $G = (V, E)$

if $\displaystyle p(x) = c_0 \cdot \prod_{(x_i,x_j)\in E} P_e(x_i, x_j) \cdot \prod_{x_i \in V} P_v(x_i)$

normalization constant

<u>Gaussians</u>

$$\log p(x) = \log \exp\{x^T \Sigma^{-1} x\} \cdot c$$

"It factors" $\displaystyle = \log c_0 + \sum_{(x_i,x_j)\in E} \log P_e(x_i, x_j) + \sum_{x_i \in V} \log P_{x_i}(x_i)$

$$A = \Sigma^{-1} \qquad = \sum_{i,j} A_{ij} x_i x_j$$

for $(i,j) \notin E$ $\quad \displaystyle \partial x_i \partial x_j \sum_{k,l} A_{kl} x_k x_l = (A_{ij} + A_{ji})$

Because Σ^{-1} is symmetric (covariance matrix)

$$A_{ij} = A_{ji} \Rightarrow 2A_{ij}.$$

if we ~~differentiate~~ factored expression?

$(i,j) \notin E \Rightarrow$ factored term must be 0.

$$\Rightarrow A_{ij} = \Sigma^{-1}_{ij} = 0$$

So Gaussians have this structure if (i,j) are not connected

$$\text{Graph} \Rightarrow \Sigma^{-1}_{ij} = 0$$

More Complex for Gaussians, (Ratner 2018) (Wainwright & Loh 2014)

Back to our Problem

$$\Sigma = \begin{array}{c|c|c} & \lambda_1 \cdots \lambda_3 & Y \\ \hline \begin{matrix}\lambda_1 \\ \vdots \\ \lambda_3\end{matrix} & \Sigma_O & \\ \hline Y & & 2 \end{array}$$

We see Σ_O

Assume we know Graph Structure

$\Rightarrow$ Zeros in Σ^{-1}

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

$$(\Sigma^{-1})_O = (\Sigma_O - uu^T)^{-1}$$

OBSERVE (Σ_O)

some rank one difference (matrix inverse)

$$(B - uu^T)^{-1} = B^{-1} + \frac{B^{-1}uu^TB^{-T}}{1 - u^TB^{-1}u} \qquad z = \frac{B^{-1}u}{\sqrt{1-u^TB^{-1}u}}$$

$$= \Sigma_O^{-1} + \underset{\sim\sim}{zz^T} \quad \text{rank one}$$

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

$$0 = (\Sigma_O^{-1})_{ij} + z_i z_j$$

so for every missing edge we get an equation.

$(\Sigma_O^{-1})_{ij} = B_{ij}$

$\boxed{B_{ij}^2 = z_i^2 z_j^2} \mapsto \log B_{ij}^2 = \log z_i^2 + \log z_j^2$

LINEAR

STRUCTURAL ABOUT INVERSE COVARIANCE TO CREATE A SEQUENCE of linear EQUATIONS → Solved those.

⇒ Solution to the correlated CASE

IN THE NOTES

- Higher RANK Correlations
- HOW CAN WE LEARN (STRUCTURE LEARN)

RECAP:

- WEAK Supervision & Augmentation
- formal theory.
- Nugget Graphs & Prob distribution
- "METHOD of MOMENTS" STYLE of Algorithms