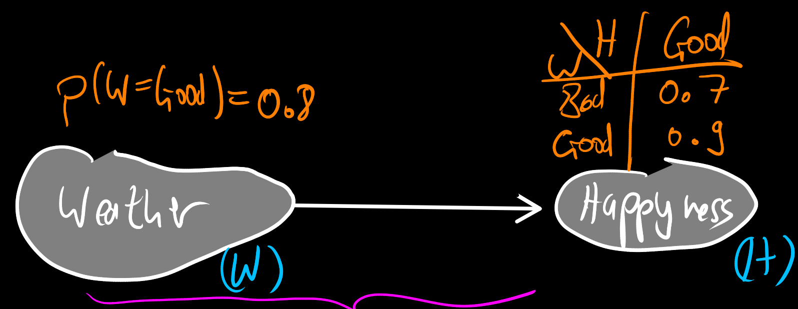
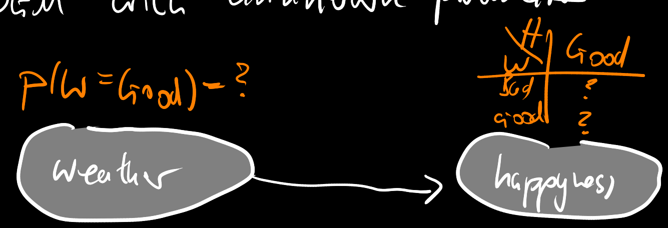
Directed Graphical Models
Maximum Likelihood Fit of Parameters



$W \in \{\overset{0}{Bad}, \overset{1}{Good}\}$

$H \in \{\overset{0}{Bad}, \overset{1}{Good}\}$

DGM with unknown parameters



we have data from it
but not the parameters

$$\mathcal{D} = \{ (w^{[0]}, h^{[0]}), \dots, (w^{[N-1]}, h^{[N-1]}) \} \quad (\text{i.i.d.})$$

e.g. (0, 1), (1, 0), ...

Likelihood:

$$\mathcal{L}(\mathcal{D}) = \prod_{i=0}^{N-1} p(W = w^{[i]}, H = h^{[i]}) \quad (\text{DGM factorizes the joint})$$

$\mathcal{L}(\mathcal{D}; \theta_W, \underline{\theta}_H)$

$$= \prod_{i=0}^{N-1} p(W = w^{[i]})\, p(H = h^{[i]} \mid W = w^{[i]})$$

$$= \prod_{i=0}^{N-1} \text{Bern}(w^{[i]}; \theta_W)\, \text{Bern}(h^{[i]}; \underline{\theta}_H[w^{[i]}])$$

Log-Likelihood

$$\ell(\underset{\text{Known}}{\mathcal{D}}; \underbrace{\theta_W, \underline{\theta}_H}_{\text{unknown}}) = \log(\mathcal{L}(\mathcal{D}; \theta_W, \underline{\theta}_H))$$

$$= \sum_{i=0}^{N-1} \log \text{Bern}(w^{[i]}; \theta_W) + \log \text{Bern}(h^{[i]}; \underline{\theta}_H[w^{[i]}])$$

Maximum likelihood

$$\theta_W, \underline{\theta}_H = \underset{\substack{\theta_W \in [0,1] \\ \underline{\theta}_H \in [0,1]^2}}{\arg\max}\ \ell(\mathcal{D}; \theta_W, \underline{\theta}_H)$$

↳ using a gradient-based optimizer

Gradients:
e.g. $\frac{\partial \ell}{\partial \theta_W} = \dots$ by Automatic Differentiation