Eigenvalues

• (positive definite) $\Rightarrow$ (eigenvalues non-negative)

Let λ be an eigenvalue of the matrix K and q be the corresponding eigenvector. Then:

$$Kq = \lambda q \Rightarrow \bar{q}^T K q = \lambda \bar{q}^T q \Rightarrow \lambda = \frac{\bar{q}^T K q}{\|q\|^2} \geq 0$$

• (eigenvalues non-negative) $\Rightarrow$ (positive definite)

The (symmetric) matrix K can be decomposed as $K = Q \Lambda Q^*$ where Q contains K's eigenvectors as columns, Λ is a diagonal matrix containing the corresponding eigenvalues and Q^* is the conjugate transpose of Q.

Then, for all $c \in \mathbb{C}^m$, we have:

$$c^T K \bar{c} = c^T Q \Lambda Q^* \bar{c} = \underbrace{(c^T Q)}_{\text{Let's call that } y} \Lambda (c^T Q)^* = y \Lambda y^* = \sum_{i=1}^{m} \lambda_i \cdot y_i \cdot \bar{y}_i$$

$$= \sum_{i=1}^{m} \lambda_i \cdot |y_i|^2$$

Since all λ_i's are non-negative, we get $c^T K \bar{c} \geq 0$.

Dot products are Kernels

Let $x_1, x_2, \dots, x_m \in \mathbb{R}^n$. We want to show that, for all $c \in \mathbb{R}^m$, it holds: $\sum_{i,j=1}^{m} c_i c_j \langle x_i, x_j \rangle \geqslant 0$

We have
$$\sum_{i,j=1}^{m} c_i c_j \langle x_i, x_j \rangle = \sum_{i,j=1}^{m} \langle c_i x_i, c_j x_j \rangle$$
$$= \left\langle \sum_{i=1}^{m} c_i x_i, \sum_{j=1}^{m} c_j x_j \right\rangle$$
$$= \langle x, x \rangle, \text{ where } x = \sum_{i=1}^{m} c_i x_i .$$

By the definition of the dot product, $\langle x, x \rangle \geqslant 0$.

Positive diagonal

For the simple case where $m=1$, for all $x \in \mathcal{X}$, we get the gram matrix $G = [k(x,x)]$. Since it is positive definite, we have $c \cdot k(x,x) \cdot \bar{c} \geq 0 \;\; \forall c \in \mathbb{C}$.
Therefore $c \cdot k(x,x) \cdot \bar{c} \geq 0 \Leftrightarrow k(x,x) \cdot |c|^2 \geq 0 \Leftrightarrow k(x,x) \geq 0$.

One can also think $k(x,x) = \langle \varphi(x), \varphi(x) \rangle = \|\varphi(x)\|^2 \geq 0$.

Squared error SVM

The optimization problem takes the form:

$$\min_{w\in H, \xi\in\mathbb{R}^n, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i^2$$

$$\text{s.t.} \quad y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i \quad \forall i=1,\dots,n$$
$$\xi_i \geq 0 \quad \forall i=1,\dots,n$$

The Lagrangian takes the form:

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i^2 + \sum_{i=1}^{n}\alpha_i\left[1-\xi_i - y_i(\langle w, x_i\rangle + b)\right] - \sum_{i=1}^{n}\beta_i\xi_i$$

where $\alpha_i, \beta_i \geq 0 \quad \forall i=1,\dots,n$

For the optimality conditions, we need:

$$\nabla_w L(\cdot) = 0 \Leftrightarrow w - \sum_{i=1}^{n}\alpha_i y_i x_i = 0 \Leftrightarrow w = \sum_{i=1}^{n}\alpha_i y_i x_i$$

$$\frac{\partial L(\cdot)}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\nabla_\xi L(\cdot) = 0 \Leftrightarrow 2\frac{C}{n}\xi - \alpha - \beta = 0 \Leftrightarrow \xi = \frac{n}{2C}(\alpha+\beta)$$

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i\left[1-\xi_i - y_i(\langle w, x_i\rangle + b)\right] - \sum_{i=1}^{n} \beta_i \xi_i$$

By substituting, the Lagrangian becomes:

$$L(\alpha,\beta) = \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle + \frac{n}{4C}\sum_{i=1}^{n}(\alpha_i+\beta_i)^2 + \sum_{i=1}^{n} \alpha_i$$

$$- \frac{n}{2C}\sum_{i=1}^{n} (\alpha_i+\beta_i)^2 - \sum_{i=1}^{n} \alpha_i y_i \left\langle \sum_{j=1}^{n} \alpha_j y_j x_j, x_i\right\rangle$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n} \alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle - \frac{n}{4C}\sum_{i=1}^{n}(\alpha_i+\beta_i)^2$$

and the constraints of the dual are:

- $\alpha_i, \beta_i \geq 0 \quad \forall\, i=1,\dots,n$
- $\sum_{i=1}^{n} \alpha_i y_i = 0$

At the optimum, $\xi_i = \underbrace{\max(0, 1-y_i(\langle w, x_i\rangle + b))}_{\text{hinge loss}}$

Group error penalty

The optimization problem takes the form:

$$\min_{w\in H,\ \xi\in\mathbb{R}^n,\ b\in\mathbb{R}} \frac{1}{2}\|w\|^2 + \sum_{i=1}^{\ell} C_i \xi_i$$

$$\text{s.t. } y_i^j(\langle w, x_i^j\rangle + b) \geq 1 - \xi_i \quad \forall j = 1,\dots,m_i,\ \forall i = 1,\dots,\ell$$

$$\xi_i \geq 0 \quad \forall i = 1,\dots,\ell$$

The Lagrangian takes the form:

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{\ell} C_i\xi_i + \sum_{i=1}^{\ell}\sum_{j=1}^{m_i} \alpha_i^j\left[1-\xi_i - y_i^j(\langle w, x_i^j\rangle + b)\right] - \sum_{i=1}^{\ell}\beta_i\xi_i$$

where $\alpha_i^j, \beta_i \geq 0 \quad \forall j=1,\dots,m_i,\ \forall i=1,\dots,\ell$

For the optimality conditions, we need:

$$\nabla_w L(\cdot) = 0 \Leftrightarrow w - \sum_{i=1}^{\ell}\sum_{j=1}^{m_i}\alpha_i^j y_i^j x_i^j = 0 \Leftrightarrow w = \sum_{i=1}^{\ell}\sum_{j=1}^{m_i}\alpha_i^j y_i^j x_i^j$$

$$\frac{\partial L(\cdot)}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{\ell}\sum_{j=1}^{m_i}\alpha_i^j y_i^j = 0$$

$$\frac{\partial L(\cdot)}{\partial \xi_i} = 0 \Leftrightarrow C_i - \sum_{j=1}^{m_i}\alpha_i^j - \beta_i = 0 \Leftrightarrow \beta_i = C_i - \sum_{j=1}^{m_i}\alpha_i^j$$

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l} C_i\xi_i + \sum_{i=1}^{l}\sum_{j=1}^{m_i} \alpha_i^j\left[1-\xi_i - y_i^j(\langle w, x_i^j\rangle + b)\right] - \sum_{i=1}^{l}\beta_i\xi_i$$

By substituting, the Lagrangian becomes:

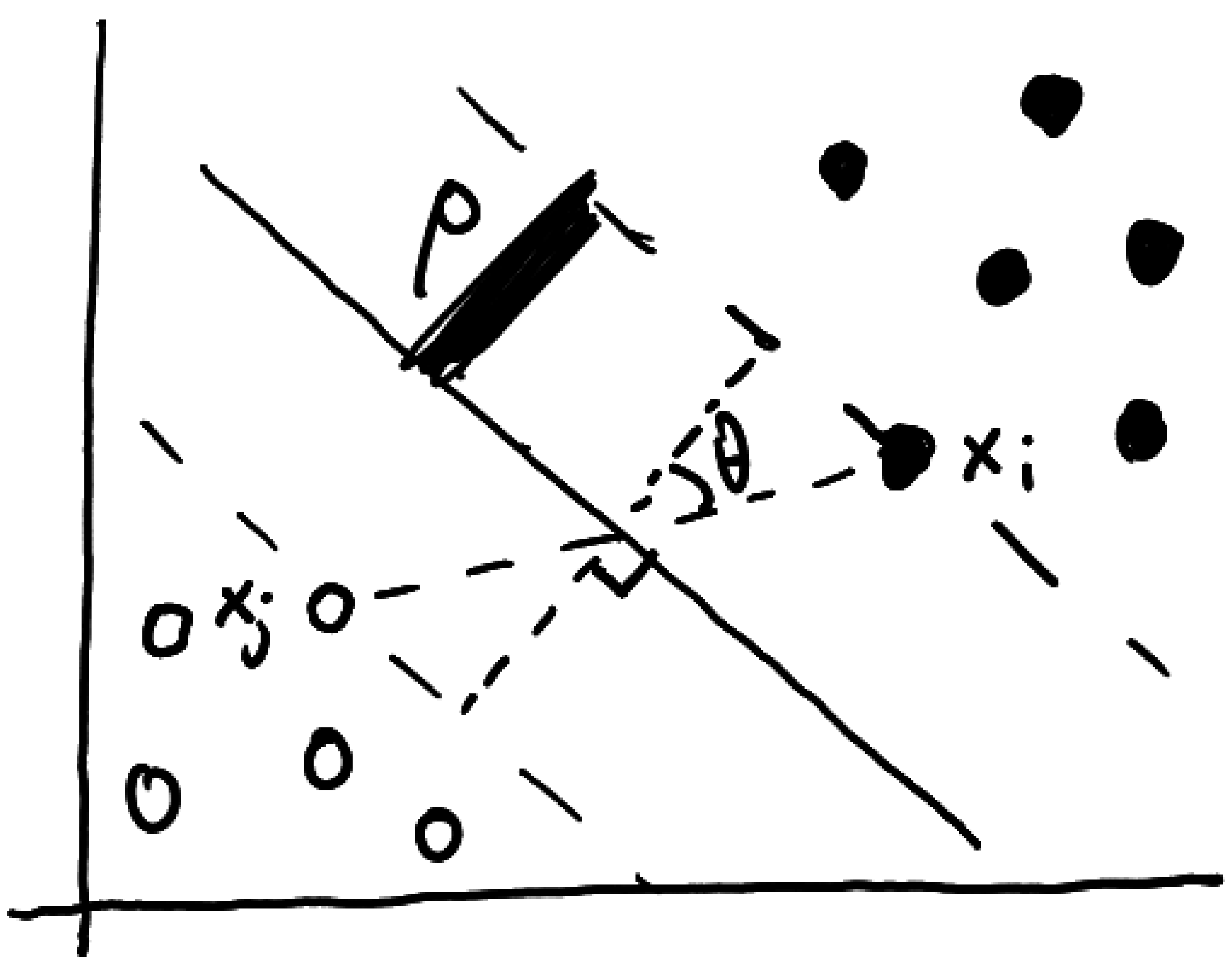
$$\begin{aligned}
L(\alpha) &= \frac{1}{2}\sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i} \alpha_i^j \alpha_{i'}^{j'} y_i^j y_{i'}^{j'} \langle x_i^j, x_{i'}^{j'}\rangle + \sum_{i=1}^{l} C_i\xi_i \\
&\quad + \sum_{i=1}^{l}\sum_{j=1}^{m_i}\alpha_i^j - \sum_{i=1}^{l}\left[\xi_i \sum_{j=1}^{m_i}\alpha_i^j\right] - \sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i} \alpha_i^j \alpha_{i'}^{j'} y_i^j y_{i'}^{j'} \langle x_i^j, x_{i'}^{j'}\rangle \\
&\quad - \sum_{i=1}^{l}\beta_i\xi_i \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m_i}\alpha_i^j - \frac{1}{2}\sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i} \alpha_i^j \alpha_{i'}^{j'} y_i^j y_{i'}^{j'} \langle x_i^j, x_{i'}^{j'}\rangle \\
&\quad + \sum_{i=1}^{l}\left[\xi_i\left(C_i - \beta_i - \sum_{j=1}^{m_i}\alpha_i^j\right)\right] \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m_i}\alpha_i^j - \frac{1}{2}\sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i} \alpha_i^j \alpha_{i'}^{j'} y_i^j y_{i'}^{j'} \langle x_i^j, x_{i'}^{j'}\rangle
\end{aligned}$$

with constraints:

- $\sum_{i=1}^{l}\sum_{j=1}^{m_i}\alpha_i^j y_i^j = 0$
- $\alpha_i^j \geq 0 \quad \forall j = 1,\dots,m_i \,.\, \forall i = 1,\dots,l$
- $\sum_{j=1}^{m_i}\alpha_i^j \leq C_i \quad \forall i = 1,\dots,l$

Margin

Let w, b be the parameters of the resulting hyperplane by solving the dual SVM problem.



For the support vectors x_i, x_j we have:

$$\left.\begin{aligned} \langle w, x_i \rangle + b &= 1 \\ \langle w, x_j \rangle + b &= -1 \end{aligned}\right\} \langle w, x_i - x_j \rangle = 2.$$

From the figure, we see that

$$\rho = \frac{1}{2}\|x_i - x_j\| \cdot \cos\theta = \frac{1}{2\|w\|} \cdot \|w\| \cdot \|x_i - x_j\| \cdot \cos\theta = \frac{1}{2\|w\|} \cdot \langle w, x_i - x_j \rangle = \frac{1}{\|w\|}.$$

If the data are linearly separable, Slater's condition is fulfilled and strong duality holds. Therefore, the optimal objective values of the primal and the dual problems match.

$$\tilde{L}(\alpha) = (\text{opt. of primal}) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \alpha_i \left[1 - y_i(\langle w, x_i \rangle + b)\right]$$

Due to the complementary slackness condition, the second term has to be equal to zero. Therefore:

$$\tilde{L}(\alpha) = \frac{1}{2}\|w\|^2 \Rightarrow \tilde{L}(\alpha) = \frac{1}{2}\frac{1}{\rho^2} \Rightarrow \frac{1}{\rho^2} = 2\tilde{L}(\alpha).$$

Moreover, we have $\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$ and, from the optimality conditions for the primal, we know that $w = \sum_{i=1}^{n} \alpha_i y_i x_i$.

Therefore:

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \|w\|^2 \Leftrightarrow$$

$$\frac{1}{2}\|w\|^2 = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\|w\|^2 \Leftrightarrow$$

$$\|w\|^2 = \sum_{i=1}^{n} \alpha_i \Leftrightarrow$$

$$\frac{1}{\rho^2} = \sum_{i=1}^{n} \alpha_i$$

SCMs

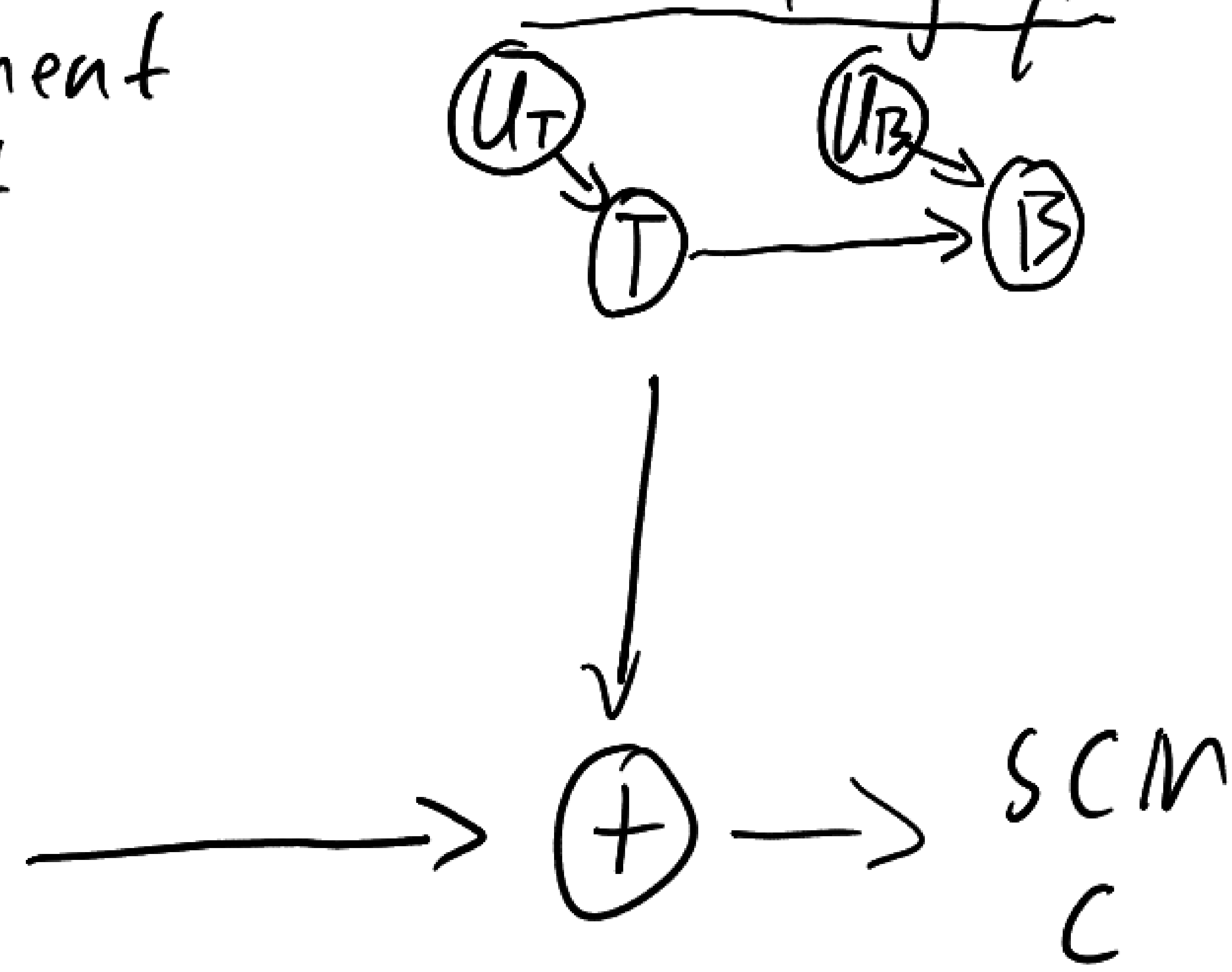
Observational, Interventional & Counterfactual Distributions

A robot tosses a coin. With heads ($U_T=1$), it treats a patient ($T=1$).

A few patients have a rare condition. ($U_B=1$)

A patient goes blind ($B=1$):
- without condition and no treatment
- with condition and treatment

Causal graph



Structural Equations

$T := U_T$

$B := T \cdot U_B + (1-T)\cdot(1-U_B)$

$U_T \sim \mathrm{Ber}(0.5)$

$U_B \sim \mathrm{Ber}(0.01)$

Observational distribution

$$P^C(B=1) = P\big((U_B=0 \wedge U_T=0) \vee (U_B=1 \wedge U_T=1)\big)$$
$$= 0.99 \times 0.5 + 0.01 \times 0.5 = 0.5$$

Interventional distribution

$$P^{C;do(T:=1)}(B=1) = P(U_B=1) = 0.01$$

Counterfactual distribution

we observe $T=B=1$. Therefore, we can infer that $U_B=1$.

Modified equations: $T := 1$, $B := T$ $\quad P^{C\mid T=B=1;\,do(T=0)}(B=0)=1$.