graph regularized matrix factorizatio

graph regularized MF

data: (compound i, compound j, miscibility outcome)
$\in \{0,1\}$ not miscible ← → miscible

$\mathbb{M}$: miscibility matrix, $\underbrace{C}_{\text{\# compounds}} \times C$, symmetric $M^T = M$.

element (i,j) of $\mathbb{M}$, m_{ij}, is miscibility outcome.

suppose only a fraction of entries are observed:

$\Omega_{obs} \subset \{1, \dots, c\} \times \{1, \dots, c\}$

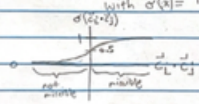
also, each compound i has class label given by $\phi(i)$,
protein, surfactant, etc.

model: $\vec{c}_i \in \mathbb{R}^d$ low dimensional latent representation for compound i

we wish to learn these...

$$\boxed{m_{ij} = \sigma(\vec{c}_i \cdot \vec{c}_j)}$$

with $\sigma(x) = 1/(1+e^{-x})$ the sigmoid function



$\vec{c}_i \cdot \vec{c}_j = \|\vec{c}_i\| \, \|\vec{c}_j\| \cos\theta$
so if vectors point in same direction, miscible, otherwise immiscible.

note: model as of now doesn't account for class labels.

loss:
$$\ell(\vec{c}_1, \dots, \vec{c}_c) = \underbrace{-\sum_{(i,j) \in \Omega_{obs}} \left\{ m_{ij} \log[\sigma(\vec{c}_i \cdot \vec{c}_j)] + (1-m_{ij}) \log[1-\sigma(\vec{c}_i \cdot \vec{c}_j)] \right\}}_{\text{cross-entropy loss}}$$
$$+ \gamma \sum_{(i,j)} \|\vec{c}_i - \vec{c}_j\|^2 \underbrace{I(\phi(i), \phi(j))}_{\text{indicator function; do these two belong to same class?}}$$

second term is graph regularization.
encourages compounds from same class to congregate.

γ hyperparam controls penalty for classes to drift away from each other.
same as making proteins fully connected graph, → spectral embedding if $\gamma \to \infty$

$\mathbb{C} = [\vec{c}_1 \cdots \vec{c}_c]$ $d \times c$ matrix. Udell: constraint $\mathbb{C}\mathbb{C}^T = I$. why?