April 28, 2020 Kernel methods

$$-\|x_i - x_j\|_2^2 = -\|x_i\|_2^2 - \|x_j\|_2^2 + 2x_i^T x_j$$

$x_i^T x_j \approx$ how similar x_i, x_j are

node2vec
t-SNE
PCA

$k(x_i, x_j) = x_i^T x_j$ (similarity)

$k(x_i, x_j) = -\|x_i - x_j\|_2^2$

Example: squared exponential / Gaussian

$$k(x_i, x_j) = s^2 \exp\left(-\frac{1}{2}\|x_i - x_j\|_2^2 / l^2\right)$$

Example: polynomial (degree 2)

$$k(x_i, x_j) = (c + x_i^T x_j)^2$$

$$(c + y^T z)^2 = \left(c + \sum_{k=1}^{d} y_k z_k\right)^2$$

$$= c^2 + 2c \sum y_k z_k + \left(\sum y_k z_k\right)^2$$

$$= c \cdot c + \sum (\sqrt{2c}\, y_k)(\sqrt{2c}\, z_k) \quad \sum_{k=1}^{d} y_k^2 z_k^2 + \sum_{i<j} (\sqrt{2} y_i y_j)(\sqrt{2} z_i z_j)$$

$$\phi(x) = (c, \sqrt{2c}\, x_1, \ldots, \sqrt{2c}\, x_d, x_1^2, \ldots, x_d^2, \sqrt{2} x_1 x_2, \ldots, \sqrt{2} x_{d-1} x_d)$$

$$= \phi(y)^T \phi(z) \qquad \phi : x \in \mathbb{R}^d \to \phi(x) \in \mathbb{R}^{1+d+d+\binom{d}{2}}$$

$= k(y, z)$

ϕ is called a feature map

What about Gaussian?

$$k(x_i, x_j) = \exp\left(-\frac{1}{2}\|x_i - x_j\|_2^2\right) \qquad r = \|x_i - x_j\|_2$$

$$\exp\left(-\tfrac{1}{2} r^2\right)$$

$$= 1 - \frac{r^2}{2} + \frac{r^4}{8} - \frac{r^6}{48} + \cdots$$

$$k(x_i, x_j) = x_i^T x_j \qquad k(x_i, x_j) = \exp\left(-\|x_i - x_j\|_2^2\right)$$

Example: kernel PCA