Discriminative Embeddings
of Latent Variable Models for Structured Data

Hanjun Dai, Bo Dai, Le Song

presentation by
Breandan Considine
McGill University

breandan.considine@mail.mcgill.ca

March 8, 2020

What is a kernel?

A feature map transforms the input space to a feature space:

$$\varphi : \overbrace{\mathbb{R}^n}^{\text{Input space}} \rightarrow \overbrace{\mathbb{R}^m}^{\text{Feature space}} \tag{1}$$

Kernel functions generalize the notion of inner products to feature maps:

$$k(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x})^\intercal \varphi(\mathbf{y}) \tag{2}$$

Gives us $\varphi(x)^\intercal \varphi(y)$ without directly computing $\varphi(x)$ or $\varphi(y)$

What is a kernel?

Consider the univariate polynomial regression algorithm:

$$\hat{f}(\mathbf{x};\boldsymbol{\beta}) = \beta\varphi(\mathbf{x}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m = \sum_{j=0}^{m} \beta_j x^j \tag{3}$$

Where $\varphi(\mathbf{x}) = [1, x_1, x_2^2, x_3^3, \ldots, x_m^m]$. We seek $\boldsymbol{\beta}$ minimizing the error:

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} ||\mathbf{Y} - \hat{\mathbf{f}}(\mathbf{X};\boldsymbol{\beta})||^2 \tag{4}$$

We can solve for $\boldsymbol{\beta}^*$ using the normal equation or gradient descent:

$$\boldsymbol{\beta}^* = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{Y} \tag{5}$$

$$\boldsymbol{\beta}' \leftarrow \boldsymbol{\beta} - \alpha\nabla_{\boldsymbol{\beta}} ||\mathbf{Y} - \hat{\mathbf{f}}(\mathbf{X};\boldsymbol{\beta})||^2 \tag{6}$$

What happens if we have a multivariate polynomial?

$$z(x, y) = 1 + \beta_x x + \beta_y y + \beta_{xy} xy + \beta_{x^2} x^2 + \beta_{y^2} y^2 + \beta_{xy^2} xy^2 + \ldots \tag{7}$$

What is a kernel?

Consider the polynomial kernel $k(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x}^T\mathbf{y})^2$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

$$k(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x}^T\mathbf{y})^2 = (1 + x_1\, y_1 + x_2\, y_2)^2 \tag{8}$$

$$= 1 + x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 y_1 + 2x_2 y_2 + 2x_1 x_2 y_1 y_2 \tag{9}$$

This gives us the same result as computing the 6 dimensional feature map:

$$k(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x})^\intercal \varphi(\mathbf{y}) \tag{10}$$

$$= [1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2]^\intercal \begin{bmatrix} 1 \\ y_1^2 \\ y_2^2 \\ \sqrt{2}y_1 \\ \sqrt{2}y_2 \\ \sqrt{2}y_1 y_2 \end{bmatrix} \tag{11}$$

But does not require computing $\varphi(x)$ or $\varphi(y)$.

Examples of common kernels

Popular kernels

Polynomial	$k(\mathbf{x},\mathbf{y}) := (\mathbf{x}^T\mathbf{y}+r)^n$	$\mathbf{x},\mathbf{y} \in \mathbb{R}^d, n \in \mathbb{N}, r \geq 0$
Laplacian	$k(\mathbf{x},\mathbf{y}) := exp\left(-\alpha\lVert\mathbf{x}-\mathbf{y}\rVert\right)$	$\mathbf{x},\mathbf{y} \in \mathbb{R}^d, \alpha > 0$
Gaussian RBF	$k(\mathbf{x},\mathbf{y}) := \exp\left(-\frac{\lVert\mathbf{x}-\mathbf{y}\rVert^2}{2\sigma^2}\right)$	$\mathbf{x},\mathbf{y} \in \mathbb{R}^d, \sigma > 0$

Popular Graph Kernels

RW	$k_\times(G,H) := \sum_{i,j=1}^{\lvert V_\times\rvert}[\sum_{n=1}^{\infty}\lambda^n A_\times^n]_{ij} = \mathbf{e}^\intercal(\mathbf{I}-\lambda A_\times)^{-1}\mathbf{e}$	$\mathcal{O}(n^6)$
SP	$k_{SP}(G,H) := \sum_{s_1\in SD(G)}\sum_{s_2\in SD(H)} k(s_1,s_2)$	$\mathcal{O}(n^4)$
WL	$\mathrm{l}^{(i)}(G) := \begin{cases} \deg_v, \forall v \in G & \mathrm{i}=1 \\ \mathrm{HASH}(\{\{\mathrm{l}^{(i-1)}(u), \forall u \in \mathcal{N}(v)\}\}) & \mathrm{i}>1 \end{cases}$ $k_{WL}(G,H) := \langle\psi_{WL}(G), \psi_{WL}(H)\rangle$	$\mathcal{O}(hm)$

https://people.mpi-inf.mpg.de/~mehlhorn/ftp/genWLpaper.pdf

What is an inner product space?

Let X be a vector space over the reals.

Definition

A function $f : X \to \mathbb{R}$ is **linear** iff $f(\alpha x) = \alpha f(x)$ and $f(x + z) = f(x) + f(z)$ for all $\alpha \in \mathbb{R}, x, z \in X$.

Definition

X is an **inner product space** if there exists a symmetric bilinear map $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ if $\forall \mathbf{x} \in X, \langle \mathbf{x}, \mathbf{x} \rangle > 0$ (i.e. is positive definite).

Scalar Product	Vector Dot Product	Random Variable
$\langle x, y \rangle := xy$	$\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle := x^{\mathsf{T}} y$	$\langle X, Y \rangle := \mathrm{E}(XY)$

What is a Hilbert space?

Let $d : X \times X \to \mathbb{R}^{\geq 0}$ be a metric on the space X.

Definition: Cauchy sequence

A sequence $\{x_n\}$ is called a **Cauchy sequence** if
$\forall \varepsilon > 0, \exists N \in \mathbb{N},$ such that $\forall n, m \geq N, d(x_n, x_m) \leq \varepsilon$.

Definition: Completeness

X is called **complete** if every Cauchy sequence converges to a point in X.

Definition: Separability

X is called **separable** if there exists a sequence $\{x_n\}_{n=1}^{\infty} \in X$ s.t. every nonempty open subset of X contains at least one element of the sequence.

Definition: Hilbert space

A Hilbert space $\mathcal{H}$ is an inner product space that is complete and separable.

Properties of Hilbert Spaces

Hilbert space inner products are kernels

The inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a positive definite kernel:

$$\sum_{i,j=1}^{n} c_i c_j (x_i, x_j)_{\mathcal{H}} = \left(\sum_{i=1}^{n} c_i x_i, \sum_{j=1}^{n} c_j x_j \right)_{\mathcal{H}} = \left\| \sum_{i=1}^{n} c_i x_i \right\|_{\mathcal{H}}^{2} \geq 0$$

Reproducing Kernel Hilbert Space (RKHS)

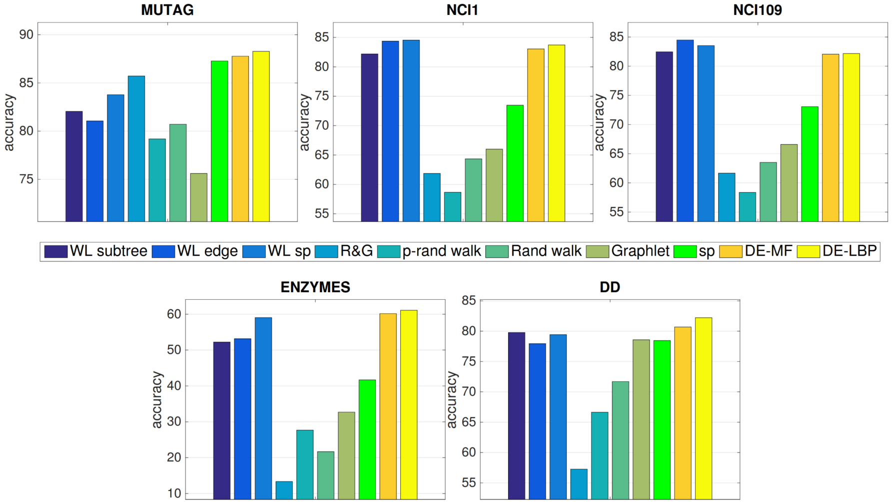
Any continuous, symmetric, positive definite kernel $k : X \times X \to \mathbb{R}$ has a corresponding Hilbert space, which induces a feature map $\varphi : X \to \mathcal{H}$ satisfying $k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$.

http://jmlr.csail.mit.edu/papers/volume11/vishwanathan10a/vishwanathan10a.pdf

Gaussian RBF kernel

Belief propagation

Results


MUTAG
NCI1
NCI109
ENZYMES
DD
accuracy
90
85
80
75
70
65
60
55
50
40
30
20
10
WL subtree
WL edge
WL sp
R&G
p-rand walk
Rand walk
Graphlet
sp
DE-MF
DE-LBP

Resources

- Properties of kernels
- Survey on Graph Kernels
- Notes Metric Spaces