Ludwig-Maximilians-Universitaet Muenchen 28.05.2021
Institute for Informatics
Prof. Dr. Volker Tresp
Rajat Koner
Andreas Lohrer
Sebastian Schmoll

Machine Learning
Summer 2021
Exercise Sheet 7

Exercise 7-1 Curse of Dimensionality vs. Kernel Trick

a) Explain the term *curse of dimensionality*.
When does it occur, how can it be avoided?

Possible Solution

(a) Curse of Dimensionality:

E.g.: 100 observations cover 1D-space ($[0, 1]$) pretty well. To gain this coverage in a 100D-space around 10^{20} samples are required, according to Leao Breiman (americ. statistician).

$\Rightarrow$ High costs in ...

- data acquisition - if possible in the first place (cf. interviewing of 100 people is not a problem, 10^{20} people don't even exist),
- data processing: $n \times n$ matrix multiplication and inversion alone has $O(n^{2.379})$ with the algorithm of Coppersmith-Winograd, distanz calculations getting very expensive due to the mass
- data storage
- Indexing hardly possible (starting at around 6-10 dimensions), since distances are hardly meaningful $\Rightarrow$ even more costs

Counteract by:

- Regularisation
- Dimension reduction (PCA, LDA, RCA)

b) Explain the term *Kernel Trick*.
How can it be used, what is its connection to the *curse of dimensionality*?

Possible Solution

b) Kernel Trick:

What: Transformation of the data from a low to a high dimensional space, e.g. by a more complex model, a diffenrent kind of feature extraction (additional features, higher granularity), etc.

Why: Separating planes can be found easier.

CoD: The part of the Curse of Dim. which is rather negative can be beneficial here. The price is the above mentioned negative characteristic.

Exercise 7-2 Kernel - Feature Mapping

In this exercise we want to compute the explicit representation of some kernels.

(a) The homogeneous quadratic kernel $K(x, y) = \langle x, y\rangle^2$ defined on the 2-dimension real vector space.

Possible Solution

$$
\begin{aligned}
K(x,y) &= \langle x, y\rangle^2 \\
&= (x_1y_1 + x_2y_2)^2 \\
&= x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \\
&= \left\langle \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}, \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{pmatrix} \right\rangle \\
&= \langle \phi(x), \phi(y)\rangle
\end{aligned}
$$

Hence, the homogeneous quadratic kernel corresponds to the inner product of a three-dimensional vector space. Or, in other words: the corresponding feature map has three coordinates.

(b) The gaussian radial basis function kernel $K(x, y) = \exp\left(-\gamma\|x - y\|^2\right)$ for $x, y \in \mathbb{R}$ and $\gamma > 0$.
Hint: Use the power series expansion of the exponential function: $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Possible Solution

$$
\begin{aligned}
K(x,y) &= exp(-\gamma(x-y)^2) = exp(-\gamma x^2 + 2xy\gamma - \gamma y^2) \\
&\underset{powerseries}{=} exp(-\gamma x^2 - \gamma y^2)(1 + \frac{2\gamma xy}{1!} + \frac{(2\gamma xy)^2}{2!} + \ldots) \\
&= exp(-\gamma x^2 - \gamma y^2)(1 + \sqrt{\frac{2\gamma}{1!}}x\sqrt{\frac{2\gamma}{1!}}y + \sqrt{\frac{(2\gamma)^2}{2!}}x^2\sqrt{\frac{(2\gamma)^2}{2!}}y^2 + \ldots) = \phi(x)^T\phi(y)
\end{aligned}
$$

,where
$\phi(x) = exp(-\gamma x^2)[1, \sqrt{\frac{2\gamma}{1!}}x, \sqrt{\frac{(2\gamma)^2}{2!}}x^2, \ldots]$
Thus, the corresponding feature map to the gaussian RBF kernel has infinite dimensionality.

Exercise 7-3 Kernel Combinations

In order to use a custom kernel $k(\mathbf{x},\mathbf{y})$ for $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, it must be shown that it is indeed a valid kernel. We can do that by expressing the explicit mapping of the implicit basis transformations: $k(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^T\phi(\mathbf{y})$. Another popular method of showing the validity of a kernel is representing a kernel, $k(\mathbf{x},\mathbf{y}) = k_1(\mathbf{x},\mathbf{y}) \circ k_2(\mathbf{x},\mathbf{y})$, as a combination of valid kernels combined through valid basis operations.

Show that for a valid kernel $k_l(\mathbf{x},\mathbf{y})$, where $l \in \mathbb{N}_+$, the following combinations are valid:

[label=)]**Scaling**: For $a > 0$: $k(\mathbf{x},\mathbf{y}) := a \cdot k_1(\mathbf{x},\mathbf{y})$ **Sum**: $k(\mathbf{x},\mathbf{y}) := k_1(\mathbf{x},\mathbf{y}) + k_2(\mathbf{x},\mathbf{y})$ **Linear combination**: For $w \in \mathbb{R}^d_+$: $k(\mathbf{x},\mathbf{y}) := \sum_{l=1}^{d} w_l \cdot k_l(\mathbf{x},\mathbf{y})$ **Product**: $k(\mathbf{x},\mathbf{y}) := k_1(\mathbf{x},\mathbf{y}) \cdot k_2(\mathbf{x},\mathbf{y})$ **Power**: For a $p \in \mathbb{N}_+ : k(\mathbf{x},\mathbf{y}) := (k_1(\mathbf{x},\mathbf{y}))^p$

(a) Possible Solution

a)

$$ak_1(\mathbf{x}_i,\mathbf{x}_j) = a\phi_1(\mathbf{x}_i)^T\phi_1(\mathbf{x}_j) = a\sum_{m=1}^{M}\phi_1(\mathbf{x}_i)_m\phi_1(\mathbf{x}_j)_m = \sum_{m=1}^{M}\sqrt{a}\phi_1(\mathbf{x}_i)_m\sqrt{a}\phi_1(\mathbf{x}_j)_m =$$
$$= \sum_{m=1}^{M}\phi_1'(\mathbf{x}_i)_m\phi_1'(\mathbf{x}_j)_m = \phi_1'(\mathbf{x}_i)^T\phi_1'(\mathbf{x}_j)$$

mit $\phi_1'(\mathbf{x}_i) = \sqrt{a}\phi_1(\mathbf{x}_i)$

b)

$$k_1(\mathbf{x}_i,\mathbf{x}_j) + k_2(\mathbf{x}_i,\mathbf{x}_j) = \phi_1(\mathbf{x}_i)^T\phi_1(\mathbf{x}_j) + \phi_2(\mathbf{x}_i)^T\phi_2(\mathbf{x}_j) = \phi_{12}''(\mathbf{x}_i)^T\phi_{12}''(\mathbf{x}_j)$$

with $\phi_{12}''(\mathbf{x}_i) = \begin{pmatrix} \phi_1(\mathbf{x}_i) \\ \phi_2(\mathbf{x}_i) \end{pmatrix}$

Possible Solution

c)

$$\sum_{l=1}^{n} w_l k_l(\mathbf{x}_i,\mathbf{x}_j) \overset{a)}{=} \sum_{l=1}^{n}\phi_l'(\mathbf{x}_i)\phi_l'(\mathbf{x}_j) \overset{b)}{=} \phi_{1\ldots l}''(\mathbf{x}_i)^T\phi_{1\ldots l}''(\mathbf{x}_j)$$

with $\phi_l'(\mathbf{x}_i) = \sqrt{w_l}\phi_l(\mathbf{x}_i)$ and $\phi_{1\ldots n}''(\mathbf{x}_i) = \begin{pmatrix} \phi_1'(\mathbf{x}_i) \\ \phi_2'(\mathbf{x}_i) \\ \vdots \\ \phi_n'(\mathbf{x}_i) \end{pmatrix}$

d)

$$
\begin{aligned}
k_1(\mathbf{x}_i,\mathbf{x}_j)\cdot k_2(\mathbf{x}_i,\mathbf{x}_j) &= \left(\sum_{m=1}^{M_1}\phi_1(\mathbf{x}_i)_m\cdot\phi_1(\mathbf{x}_j)_m\right)\cdot\left(\sum_{n=1}^{M_2}\phi_2(\mathbf{x}_i)_n\cdot\phi_2(\mathbf{x}_j)_n\right)=\\
&=\sum_{m=1}^{M_1}\left(\phi_1(\mathbf{x}_i)_m\phi_1(\mathbf{x}_j)_m\cdot\left(\sum_{n=1}^{M_2}\phi_2(\mathbf{x}_i)_n\phi_2(\mathbf{x}_j)_n\right)\right)=\\
&=\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}(\phi_1(\mathbf{x}_i)_m\phi_1(\mathbf{x}_j)_m\cdot\phi_2(\mathbf{x}_i)_n\phi_2(\mathbf{x}_j)_n)=\\
&=\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}(\phi_1(\mathbf{x}_i)_m\phi_2(\mathbf{x}_i)_n)\cdot(\phi_1(\mathbf{x}_j)_m\phi_2(\mathbf{x}_j)_n)=\\
&=\phi_{12}'''(\mathbf{x}_i)^T\phi_{12}'''(\mathbf{x}_j)=k(x_i,x_j)
\end{aligned}
$$

$$
\text{mit } \phi_{12}'''(\mathbf{x}_i)=\begin{pmatrix}\phi_1(\mathbf{x}_i)_1\cdot\phi_2(\mathbf{x}_i)_1\\ \phi_1(\mathbf{x}_i)_1\cdot\phi_2(\mathbf{x}_i)_2\\ \vdots\\ \phi_1(\mathbf{x}_i)_1\cdot\phi_2(\mathbf{x}_i)_{M_2}\\ \phi_1(\mathbf{x}_i)_2\cdot\phi_2(\mathbf{x}_i)_1\\ \vdots\\ \phi_1(\mathbf{x}_i)_{M_1}\cdot\phi_2(\mathbf{x}_i)_{M_2}\end{pmatrix}
$$

e) Proof by induction:

I.H.: $k(\mathbf{x}_i,\mathbf{x}_j) := k_1(\mathbf{x}_i,\mathbf{x}_j)^p$ is a kernel (for p positive integer).

I.B.: $k_1(x_i,x_j)\cdot k_1(x_i,x_j)$ is a valid kernel. Proofen in d).

I.S.: $k_1(\mathbf{x}_i,\mathbf{x}_j)^{p+1} = k_1(\mathbf{x}_i,\mathbf{x}_j)^p\cdot k_1(\mathbf{x}_i,\mathbf{x}_j) \overset{\text{I.H.}}{=} \text{Kern}\cdot k_1(\mathbf{x}_i,\mathbf{x}_j) \overset{\text{d)}}{=}$ valid kernel.