REGRESSION II

Machine Learning for Autonomous Robots

Bilal Wehbe
RIC

DFKI Bremen

30 novembre 2023 – Bremen, Deutschland

University of Bremen

German Research Center for Artificial Intelligence

Linear Algebra Recap

Def. : Group, Abelian Group

Let G be a set and $\circ : G \times G \to G$ a binary operation on that set. $(G, \circ)$ is called a group iff $\forall a, b, c \in G$:

- $a \circ b \in G$ (Closure).
- $(a \circ b) \circ c = a \circ (b \circ c)$ (Associativity).
- $\exists e \in G : a \circ e = e \circ a = a$ (Identity).
- $\forall a, \exists -a : a \circ -a = -a \circ a = e$ (Inverse).

If furthermore $a \circ b = b \circ a$ (Commutativity) the group is called **Abelian**. **Example :** $(\mathbb{Z}, +)$.

Def. : Field

Let K be a set, $+, *$ two binary operations on K and $0, 1$ distinct elements in K. $(K, +, *)$ is called a **field** iff :

- $(K, +)$ is an abelian group with neutral element 0,
- $(K \setminus \{0\}, *)$ is an abelian group with neutral element 1 and
- The distributive law

$$
\begin{aligned}
a * (b + c) &= a * b + a * c \\
(a + b) * c &= a * c + b * c
\end{aligned}
$$

  applies for all $a, b, c \in K$.

Example : $\mathbb{R}$ and $\mathbb{C}$.

Def. : Vector Space

$\mathcal{X}$ is called **vector space** over field $\mathbb{F}$ iff for all $x, y, z \in \mathcal{X}$ and $a, b \in \mathbb{F}$

- $(\mathcal{X}, +)$ is a group.
- $ax \in \mathcal{X}$ (Closure).
- $a(x + y) = ax + ay$ and $(a + b)x = ax + bx$ (Distributivity).
- $a(bx) = (ab)x$ (Multiplicative Associativity).
- $1x = x$ (Multiplicative Identity).

Example : For all $d \in \mathbb{N}$, $\mathbb{R}^d$ is a vector space over $\mathbb{R}$.

Def. : Inner Product Space

Let $\mathcal{X}$ be a vector space over a field $\mathbb{F}$ with an inner product $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfying the following conditions :

- $\langle x, y \rangle = \langle y, x \rangle$ (Symmetry).
- $a\langle x, y \rangle = \langle ax, y \rangle$ and $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ (Linearity).
- $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \Leftrightarrow x = 0$ (Positive semi-definiteness).

for all $x, y, z \in \mathcal{X}$ and all $a \in \mathbb{F}$. A vectot space with inner product is called **inner product space**.

Kernel Functions

Def. : Kernel Function

Let $\mathcal{X}$ be a vector space and $\mathcal{F}$ an inner product space. A function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a **Kernel Function** or **Kernel** iff for all $x, y \in \mathcal{X}$:

$$\begin{aligned}\kappa(x, y) &= \langle \phi(x), \phi(y) \rangle_{\mathcal{F}} \\ &= \langle \phi(y), \phi(x) \rangle_{\mathcal{F}} \\ &= \kappa(y, x)\end{aligned}$$

for a **Feature Mapping** $\phi : \mathcal{X} \to \mathcal{F}$. $\mathcal{F}$ is then also called **Feature Space**.

- To prove, that some κ is a Kernel Function, it is sufficient to show, that such a ϕ exists.
- This is usually hard, it is common, that neither ϕ not $\mathcal{F}$ are explicitly known.
- We will now take a look at some proofs related to kernel functions. They are mostly based on **Construction** from the linear kernel.
- Alternatively we can prove, that the **Kernel-Matrix** of a kernel is positive semi-definite (we get to that).

First some tools

Def. : Kernel Matrix, Gram Matrix

The **Kernel Matrix** or **Gram Matrix** K is created, by applying a kernel κ to all pairs of vectors in a set $\{x_i\}_{i=1}^{n}$:

$$K := \begin{pmatrix} \kappa(x_1, x_1) & \kappa(x_1, x_2) & \cdots & \kappa(x_1, x_n) \\ \kappa(x_2, x_1) & \kappa(x_2, x_2) & \cdots & \kappa(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(x_n, x_1) & \kappa(x_n, x_2) & \cdots & \kappa(x_n, x_n) \end{pmatrix}$$

Def. : Positive-definiteness for Matrices

Let M be a $n \times n$ square matrix. M is called positve definite iff :

$$x^\intercal M x > 0$$

and positve semi-definite iff :

$$x^\intercal M x \geq 0$$

for all n-vectors x. For a positve (semi-) definite Matrix all Eigenvalues are > 0 (≥ 0).

Prop. :

All Kernel Matrices are positve semi-definite.

Proof : Let K be a kernel matrix for any dataset $\{x_i\}_{i=1}^n$ and kernel κ. Then for any $y \in \mathbb{R}^n$ it holds :

$$\begin{aligned} y^\intercal K y &= \sum_{i,j} y_i \kappa(x_i, x_j) y_j \\ &= \sum_{i,j} y_i \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{F}} y_j \\ &= \sum_{i,j} \langle y_i \phi(x_i), y_j \phi(x_j) \rangle_{\mathcal{F}} \\ &= \left\langle \sum_i y_i \phi(x_i), \sum_j y_j \phi(x_j) \right\rangle_{\mathcal{F}} \\ &\geq 0 \end{aligned}$$

$\square$

Proposition :

For any vector space $\mathcal{X}$ which is also an inner product space, the **linear kernel** $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$\kappa(x, y) := \langle x, y \rangle$$

Proof : If $\mathcal{X}$ is an inner product space, κ obviously suffices the kernel properties. $\square$

The feature map is the identity map $\phi(x) = x$, the feature space is $\mathcal{X}$ itself.

Proposition :

Let κ be a kernel, then for all $a \in \mathbb{R}, a > 0$

$$\omega := a\kappa$$

is also a kernel.

Proof : Let K be a kernel matrix for any dataset $\{x_i\}_{i=1}^n$ and kernel κ, and O accordingly for kernel ω. Then for any $y \in \mathbb{R}^n$ it holds :

$$\begin{aligned} \omega = a\kappa &\Rightarrow O = aK \\ &\Rightarrow y^\intercal O y = a y^\intercal K y \geq 0 \end{aligned}$$

$\square$

The related feature space of ω is the feature space of κ scaled by $\sqrt{a}$.

Proposition

Let κ_1 and κ_2 be kernels. Then

$$\kappa = \kappa_1 + \kappa_2$$

is also a kernel.

Proof : Let K, K_1, K_2 be kernel matrices for $\kappa, \kappa_1, \kappa_2$ for a dataset of size n and $y \in \mathbb{R}^n$, then :

$$\begin{aligned} \kappa = \kappa_1 + \kappa_2 &\Rightarrow K = K_1 + K_2 \\ &\Rightarrow y^\intercal K y = y^\intercal K_1 y + y^\intercal K_2 y \geq 0 \end{aligned}$$

$\square$

The related feature space of κ is then $\begin{pmatrix} \phi(x) \end{pmatrix} = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \end{pmatrix}$.

Def. : Spectral Decomposition

Any quadratic and symmetric $n \times n$ matrix M can be written as :

$$M = \sum_{i=1}^{n} \lambda_i v_i v_i^\intercal$$

where λ_i are the Eigenvalues and v_i the corresponding Eigenvectors of M. This is called the **Spectral Decomposition** of M.

Proof : We observe :

- The eigenvectors for different eigenvalues are linearly independent, therefore any $n \times n$ matrix N satisfying $Nv_i = Mv_i$ for all $i \in \{1, \ldots, n\}$ is necessarily indetical to M.
- Eigenvectors for different eigenvalues are orthogonal, therefore $\langle v_i, v_j \rangle = 0$ for all $i \neq j$.

Therefore for any j

$$\left(\sum_{i=1}^{n} \lambda_i v_i v_i^\intercal\right) v_j = \sum_{i=1}^{n} \lambda_i v_i v_i^\intercal v_j = \lambda_j v_j = M v_j$$

and therefore

$$\sum_{i=1}^{n} \lambda_i v_i v_i^\intercal = M$$

$\square$

Proposition

Let κ_1 and κ_2 be kernels. Then

$$\kappa = \kappa_1 \cdot \kappa_2$$

is also a kernel.

Proof : Let K, K_1, K_2 be kernel matrices for $\kappa, \kappa_1, \kappa_2$ for a dataset of size n and $y \in \mathbb{R}^n$. $\odot$ is the element-wise product for matrices :

$$\begin{aligned}
\kappa = \kappa_1 \cdot \kappa_2 \Rightarrow K &= K_1 \odot K_2 \\
&= \sum_{i=1}^{n} \lambda_i v_i v_i^{\mathsf{T}} \odot \sum_{j=1}^{n} \gamma_j w_j w_j^{\mathsf{T}} \\
&= \sum_{i,j=1}^{n} \sqrt{\lambda_i \gamma_j} (v_i \odot w_j)(v_i \odot w_j)^{\mathsf{T}}
\end{aligned}$$

Which is again a positive semi-definite matrix.

The related feature space : $\big(\phi(x)\big)_{ij} = \big(\phi_1(x)\big)_i \cdot \big(\phi_2(x)\big)_j$.

Lemma

$$\sum_{i,j=1}^{n} \sqrt{\lambda_i \gamma_j} (v_i \odot w_j)(v_i \odot w_j)^\intercal$$

is a positive semi-definite matrix.

Proof : For any $n \times n$ matrix of the form $M = \sum_i x_i x_i^\intercal$ with $x_i \in \mathbb{R}^n$ for all $y \in \mathbb{R}^n$

$$\begin{aligned} y^\intercal M y &= y^\intercal \left(\sum_i x_i x_i^\intercal \right) y \\ &= \sum_i (y^\intercal x_i)^2 \\ &\geq 0 \end{aligned}$$

□

Proposition :

Let κ be a kernel and $p : \mathbb{R} \to \mathbb{R}$ a polynomial with non-negative coefficients.

$$\omega := p(\kappa)$$

is also a kernel.

Proof : Follows from the previous oberservations of the sum, product and scalar product for kernels. $\square$

Proposition

Let κ be a kernel.

$$\omega = e^{\kappa}$$

is also a kernel.

Proof : Part of the exercise sheet !

Thank You !