TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

X Math Crash-Course for Machine Learning 1

Machine Learning 1: Foundations

Billy Joe Franks *(TUK)*

Can you follow?

$$\underset{b\in\mathbb{R},\boldsymbol{\alpha}\in\mathbb{R}^d}{\arg\min}\ \frac{1}{2}\boldsymbol{\alpha}^T K\boldsymbol{\alpha} + c\mathbf{1}^T\hat{\mathrm{l}}(\mathbf{y}\circ K\boldsymbol{\alpha} + b\mathbf{1})$$

$$K := XX^T$$

$$\mathrm{l}(x) := \max(0, 1 - x)$$

Math used in Machine Learning

Machine Learning mostly requires

- Linear Algebra
- Multivariate Calculus
- Optimization

Linear Algebra & Analysis

We will recap the following topics in Linear Algebra and Analysis

- Vectors & Matrices
- Scalar Product & Projection
- Dimension Theorem
- Eigenvalues & Eigenvectors
- Matrix Decompositions
- Gradient
- Jacobian & Hessian Matrix

Vectors

In ML1 we represent vectors as $\mathbf{v} \in \mathbb{R}^d$.

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}$$

The transpose of a vector is

$$\mathbf{v}^\top := \begin{pmatrix} v_1, \dots, v_d \end{pmatrix}$$

Matrices

Similarly for a matrix $A \in \mathbb{R}^{m \times n}$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, A^\top := \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & & \ddots & \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

We denote with $A_{i\cdot}$ the i-th row of the matrix and with $A_{\cdot i}$ the i-th column of the matrix.

Scalar Product

The scalar product for $\mathbf{v},\mathbf{w} \in \mathbb{R}^d$ is defined as:

$$\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^\top \mathbf{w} := \sum_{i=1}^{d} v_i w_i.$$

The scalar product is a bilinear form, it fulfills the following properties.

- $\langle v_1 + v_2, w \rangle = \langle v_1, w \rangle + \langle v_2, w \rangle$
- $\langle v, w_1 + w_2 \rangle = \langle v, w_1 \rangle + \langle v, w_2 \rangle$,
- $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$
- $\langle v, w\lambda \rangle = \langle v, w \rangle \lambda$.

Proof is left as an exercise.

Projection

Let $||\mathbf{v}|| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^\top \mathbf{v}}$ be the norm of a vector.

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d \backslash \{0\}$. The scalar projection of $\mathbf{v}$ onto $\mathbf{w}$ is defined as $\Pi_w(v) := \dfrac{\mathbf{w}^\top \mathbf{v}}{\|\mathbf{w}\|}$.

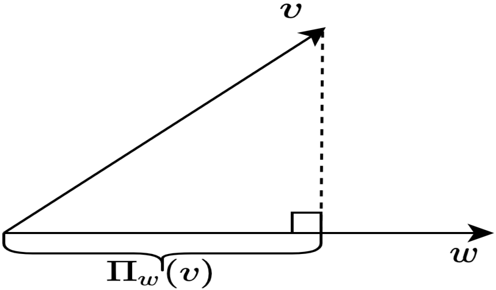

Figure: Geometrical illustration of the scalar projection. Given the vectors $\mathbf{v}$ and $\mathbf{w}$, here $\Pi_w(v)$ is the scalar projection of $\mathbf{v}$ onto $\mathbf{w}$.

Matrix Multiplication

For the following matrices $C = AB$ if

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix}$$

then $\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{pmatrix}$ where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, p$.

Matrices as Functions

A Matrix $A \in \mathbb{R}^{m \times n}$ defines a function

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}^m \quad \text{with} \quad v \mapsto Av$$

In particular this function is linear as:

- $A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w}$ (vector addition)
- $\lambda A\mathbf{v} = A\lambda\mathbf{v}$ (scalar multiplication)

Proof left as an exercise. Define two important sets:

$$Ker(A) := \{\mathbf{v} : A\mathbf{v} = \mathbf{0}\} \subseteq \mathbb{R}^n$$

$$Im(A) := \{\mathbf{w} : A\mathbf{v} = w\} \subseteq \mathbb{R}^m$$

Intuitively, multiplying by a matrix transforms a vector to another one. Important to note that this transformation is linear. It could be a rotation or for example a stretching.

Vector Space

Let $W \subseteq \mathbb{R}^d$. W is a vector space iff

- $\forall \mathbf{v}, \mathbf{w} \in W : \mathbf{v} + \mathbf{w} \in W$
- $\forall \mathbf{w} \in W, \lambda \in \mathbb{R} : \lambda \mathbf{w} \in W$

Any set $X \subseteq \mathbb{R}^d$ can be extended to a vector space and we call this extension the span $span(X)$.

Dimension

Let $W \subseteq \mathbb{R}^d$ be a vector space. We call $dim(W)$ the minimal number n such that there exist vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ such that for any $\mathbf{w} \in W$ there exist $\lambda_1, \lambda_2 \ldots \lambda_n$ with

$$\mathbf{w} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n.$$

If this holds then $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ form a basis of W. Such a basis always exists and vectors can be removed or added to complete a basis.

Dimension Theorem

Let $V \subseteq \mathbb{R}^d$ be a vector space and $f : V \to \mathbb{R}^q$ be a linear function then the following holds:

$$\dim V = \dim\left(\ker(f)\right) + \dim(\mathrm{im}(f))$$

Proof:

- Let B be a basis of ker(f).
- Complete B with A to be a basis of V where $A \cap B = \{\}$
- $\hat{f}(A) = \{f(a) \mid a \in A\}$ is a basis of im(f) as the restriction $f' : span(A) \to f(span(A))$ of f onto $span(A)$ is injective ($ker(f') = \mathbf{0}$) clearly surjective and as $f(B) = \mathbf{0}$
- $\dim V = |A| + |B| = \dim(\mathrm{im}(f)) + \dim\left(\ker(f)\right)$