Machine Learning I: Foundations
Exercise Sheet 1

Prof. Marius Kloft TA: Billy Joe Franks

17.05.2021
Deadline: 11.05.2021

1) (MANDATORY) 10 Points

In this question we will be exploring eigenvectors and eigenvalues. Let $A \in \mathbb{R}^{d\times d}$. Recall the following definition from linear algebra: a vector $\mathbf{v} \in \mathbb{R}^d$ is an *eigenvector* of A if and only if there exists $\lambda \in \mathbb{R}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. We then call λ the *eigenvalue* of A associated with the vector $\mathbf{v}$. Note that, if $\mathbf{v}$ is an eigenvector of A, then, for any $a \in \mathbb{R}$, $a\mathbf{v}$ is also an eigenvector of A. Therefore, $\mathbf{v}$ and $a\mathbf{v}$ are not considered 'distinct' eigenvectors. Prove the following:

Proposition 1 *If A has a finite number of distinct eigenvectors then each eigenvector must have a unique eigenvalue.*

Proposition 1 can be decomposed in the two propositions:
P: A has a finite number of eigenvectors
Q: Each eingenvector has a distinct eigenvalue

From the material conditional, we can extract that $P \to Q \equiv \neg Q \to \neg P$. So we will proceed by contrapositive and thus prove that: if each eigenvector does not have a distinct eigenvalue then A has (must have) an infinite number of distinct eigenvectors.

Since each eigenvector does not have a distinct eigenvalue, there must be two distinct eigenvectors, $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that

$$\begin{aligned} \lambda\mathbf{x} &= A\mathbf{x} \\ \lambda\mathbf{x}' &= A\mathbf{x}'. \end{aligned}$$

Let $\alpha, \beta \in \mathbb{R}$ be nonzero. From this we have

$$\begin{aligned} A(\alpha\mathbf{x} + \beta\mathbf{x}') &= \alpha A\mathbf{x} + \beta A\mathbf{x}' \\ &= \alpha\lambda\mathbf{x} + \beta\lambda\mathbf{x}' \\ &= \lambda(\alpha\mathbf{x} + \beta\mathbf{x}'), \end{aligned}$$

so $\alpha\mathbf{x} + \beta\mathbf{x}'$ is also an eigenvector of A with eigenvalue λ. Using this it is possible to construct an infinite number of distinct eigenvectors by varying α and β.

2) (MANDATORY) 10 Points

Recall the following definition from linear Algebra: a symmetric matrix $A \in \mathbb{R}^{d\times d}$ is called positive definite, if $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^d$ with $\mathbf{x} \neq \mathbf{0}$. Let A be symmetric.

a) Prove that, if all eigenvalues of A are positive, then A is positive definite.

A symmetric real matrix A can be diagonalized by an orthogonal matrix P, that is $A = PDP^\top$, where D is a diagonal matrix that contains the eingenvalues of the matrix in its diagonal. Let $\mathbf{x}' = P^\top\mathbf{x}$,

$$\begin{aligned} \mathbf{x}^\top A\mathbf{x} &= \mathbf{x}^\top PDP^\top\mathbf{x} \\ &= \mathbf{x}'^\top D\mathbf{x}' \qquad\qquad \text{NOTE: } \mathbf{x}'^\top = \mathbf{x}^\top P. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbf{x}'^\top D\mathbf{x}' &= \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} \times \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 y_1 & \lambda_2 y_2 & \dots & \lambda_n y_n \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \lambda_1 {y_1}^2 + \lambda_2 {y_2}^2 + \dots + \lambda_n {y_n}^2. \end{aligned}$$

By the assumption that all $\lambda_i > 0$ we have $\mathbf{x}^\top A\mathbf{x} = \mathbf{x}'^\top D\mathbf{x}' > 0$. Therefore A is positive definite.(Note: P is invertible and $\mathbf{x} \neq \mathbf{0}$, so $\mathbf{x}' \neq \mathbf{0}$)

b) Prove that, all eigenvalues of A are positive, if A is positive definite.

Let $\lambda \in \mathbb{R}$ be an eigenvalue of matrix A and $\mathbf{v}$ its corresponding eigenvector. From this we have $A\mathbf{v} = \lambda\mathbf{v}$. Multiplying $\mathbf{v}^\top$ on both sides

$$\begin{aligned} \mathbf{v}^\top A\mathbf{v} &= \mathbf{v}^\top \lambda\mathbf{v} \\ &= \lambda\mathbf{v}^\top\mathbf{v} \\ &= \lambda||\mathbf{v}||^2. \end{aligned}$$

By definition, $\mathbf{v}^\top A\mathbf{v} > 0$. Since $||\mathbf{v}||^2 > 0.$, we must have λ positive.

Now let $F : \begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R} \\ (x, y) & \mapsto & x^2 + 2y^2 + 4.97. \end{array}$

c) Find the critical point of F.

$$\nabla F = \begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x \\ 4y \end{bmatrix}$$

Being a critical point implies $\nabla F = \mathbf{0}$

$$\begin{bmatrix} 2x \\ 4y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Therefore $2x = 0$ and $4y = 0$. Trivially, the critical point is $(0, 0)$.

d) Compute the Hessian matrix H of F in any point $(x, y)^\top \in \mathbb{R}^2$.

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

e) Recall from multivariate calculus that, if H is positive definite in a critical point $(x, y)^\top$, then $(x, y)^\top$ is a local minimum. Show that the critical point of F is a local minimum. **Hint:** note that H is symmetric.

H does not depend on the point it is taken in and since it is diagonal with only positive entries it is trivially positive definite.

3) Find the global minima of the following functions $f : \mathbb{R} \to \mathbb{R}$ and $g, h, i : \mathbb{R}^2 \to \mathbb{R}$.

a) $f(w) := aw^2 + bw + c$

The first and second derivatives are:

$$\frac{\partial}{\partial w} f(w) = 2aw + b$$
$$\frac{\partial^2}{\partial w^2} f(w) = 2a$$

Setting the first derivative 0 we get:

$$0 = 2aw + b$$
$$w = -\frac{b}{2a}$$

The only critical point is $w = -\frac{b}{2a}$. According to the second derivative, if $2a > 0$ then this critical point is the global minimum as f is then convex.

b) $g(\mathbf{w}) := \mathbf{w}^T A\mathbf{w} + \mathbf{b}^T\mathbf{w} + c$

The first and second derivatives are:

$$\frac{\partial}{\partial \mathbf{w}} g(\mathbf{w}) = (A + A^T)\mathbf{w} + \mathbf{b}$$
$$\frac{\partial^2}{\partial \mathbf{w}\partial \mathbf{w}^T} g(\mathbf{w}) = (A + A^T)$$

Setting the first derivative 0 we get:

$$0 = (A + A^T)\mathbf{w} + \mathbf{b}$$
$$\mathbf{w} = -(A + A^T)^{-1}\mathbf{b}$$

According to our derivation this function has a critical point at $\mathbf{w} = -(A + A^T)^{-1}\mathbf{b}$, if $(A + A^T)$ is invertible. According to the second derivative this critical point is a global minimum if $(A + A^T)$ is positive definite as g is then convex.

Note: Typicaly matrices we consider in ML are symmetric in which case $A + A^T = 2A$

c) $h(\mathbf{w}) := aw_1^2 + bw_2 + c$

The first and second derivatives are:

$$\frac{\partial}{\partial \mathbf{w}} h(\mathbf{w}) = \begin{pmatrix} 2aw_1 \\ b \end{pmatrix}$$
$$\frac{\partial^2}{\partial \mathbf{w}\partial \mathbf{w}^T} h(\mathbf{w}) = \begin{pmatrix} 2a & 0 \\ 0 & 0 \end{pmatrix}$$

Setting the first derivative 0 we get:

$$0 = 2aw_1$$
$$w_1 = 0$$
$$0 = b$$

Thus we can follow, this function has critical points if $b = 0$, in which case it has infinitely many described by $H = \{\mathbf{x} \in \mathbb{R}^2 | \begin{pmatrix} 1 & 0 \end{pmatrix} \mathbf{x} = 0\}$. The evaluations of all points in H under h are equal and as such this function does not have a global minimum (or has infinitely many depending on your definition).

d) $i(\mathbf{w}) := w_1^2 + w_2^2 + w_1^2 w_2$

The first and second derivatives are:

$$\frac{\partial}{\partial \mathbf{w}} i(\mathbf{w}) = \begin{pmatrix} 2w_1 + 2w_1 w_2 \\ 2w_2 + w_1^2 \end{pmatrix}$$
$$\frac{\partial^2}{\partial \mathbf{w} \partial \mathbf{w}^T} i(\mathbf{w}) = \begin{pmatrix} 2 + 2w_2 & 2w_1 \\ 2w_1 & 2 \end{pmatrix}$$

Setting the first derivative 0 we get:

$$\begin{aligned} 0 &= 2w_2 + w_1^2 \\ w_2 &= -\frac{w_1^2}{2} \\ 0 &= 2w_1 + 2w_1 w_2 \\ &= 2w_1 - 2w_1 \frac{w_1^2}{2} \\ &= 2w_1 - w_1^3 \\ &= w_1(2 - w_1^2) \\ &= w_1(\sqrt{2} - w_1)(\sqrt{2} + w_1) \end{aligned}$$

From this we can follow this function has 3 critical points, namely:

$$i\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0, i\begin{pmatrix} \sqrt{2} \\ -1 \end{pmatrix} = 1, i\begin{pmatrix} -\sqrt{2} \\ -1 \end{pmatrix} = 1$$

However

$$i\begin{pmatrix} 3 \\ -2 \end{pmatrix} = -5$$

Thus none of these critical points are a global minimum.

Hint: Compute the gradient of the above functions and set it to zero. Do not forget to check the necessary conditions on global minima.

4) For a matrix $X \in \mathbb{R}^{m \times n}$ let $X_{i,:} = [X_{i,1}, \ldots, X_{i,n}]$ be the i-th row vector and $X_{:,i} = [X_{1,i}, \ldots, X_{m,i}]^T$ be the i-th column vector. For $X \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{R}^{n \times q}$ show that

$$XY = [X_{i,:} Y_{:,j}]_{i,j}$$

and

$$XY = \sum_{i=1}^{n} X_{:,i} Y_{i,:}.$$

Be sure to note the orientations of the vectors, some of these are row vectors and others are column vectors.

The first equality is the defnition of matrix product in an elementwise fashion. One way to interpret the multiplication of two matrices is as a way to concisely represent every inner product of the rows of X with the colmns of Y as another grid of number. So we will just prove this by looking at the i,jth entry of XY:

$$(XY)_{i,j} = \sum_{k=1}^{n} X_{i,k} Y_{k,j}$$

and since, for some vectors x and y, we have that

$$x^T y = \sum_{i=1}^{d} x_i y_i$$

it follows that

$$\sum_{k=1}^{n} X_{i,k} Y_{k,j} = X_{i,:} Y_{:,j}.$$

Please convince yourself of every step presented here.
For the second equality we begin with the observation that for two vectors $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ we have that

$$xy^T = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

From this we have that

$$\sum_{i=1}^{n} X_{:,i} Y_{i,:} = \sum_{i=1}^{n} \begin{bmatrix} X_{1,i}Y_{i,1} & X_{1,i}Y_{i,2} & \cdots & X_{1,i}Y_{i,q} \\ X_{2,i}Y_{i,1} & X_{1,i}Y_{i,2} & \cdots & X_{2,i}Y_{i,q} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m,i}Y_{i,1} & X_{m,i}Y_{i,2} & \cdots & X_{m,i}Y_{i,q} \end{bmatrix}.$$

Looking at some entry of this, say the (k,l)th entry, we get the summation

$$\sum_{i=1}^{n} X_{k,i} Y_{i,l} = X_{k,:} Y_{:,l}$$

which is equal to the (k,l)th entry of XY by the first equality.