Research Positions in Machine Learning and Computer Vision

Alexandre Delode

June 16, 2024

1 Question 1

Only two equations are supposed to be known:

$$\text{Area (square of side } a) = a \text{ x } a$$
$$\text{SumIntegers (} n) = \frac{n * (n+1)}{2}$$

Figure 1 illustrates how a isosceles right triangle of side a can be approximated by a stack of smaller squares with side $\frac{a}{2}$. Each row of the stack contains one more small square than the row below. Figure 2 illustrates how the stack approximates the isosceles right triangle by letting n tend to infinity. The area covered by the red smaller squares is :

$$\begin{aligned} \text{Area (square of side } \frac{a}{n}) * \text{SumIntegers } (n) &= \frac{a}{n} * \frac{a}{n} * \frac{n * (n+1)}{2} \\ &= \frac{a^2}{2}(1 + \frac{1}{n}) \\ &\xrightarrow[n \to \infty]{} \frac{a^2}{2} \end{aligned}$$

2 Question 2

The singular value decomposition theorem states that if $\mathbf{A}$ is a real $m \times n$ matrix, then there exist orthogonal matrices $\mathbf{U} = [\mathbf{u}_1 | \dots | \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ and $\mathbf{V} = [\mathbf{v}_1 | \dots | \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ such that:

$$\mathbf{U}^T \mathbf{A} \mathbf{V} = \mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad p = \min\{m, n\},$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

By multiplying by blocs, $\mathbf{A}$ can be written as a sum of rank 1 matrix :

$$\begin{aligned} \mathbf{A} &= \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \\ &= [\sigma_1 \mathbf{u}_1 | \dots | \sigma_p \mathbf{u}_p] * [\mathbf{v}_1 | \dots | \mathbf{v}_p]^T \\ &= \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \end{aligned}$$

If $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m < n, \text{rank}(\mathbf{A}) = m,$ and $\mathbf{b} \in \mathbb{R}^m,$ then the linear system $\mathbf{A}\mathbf{x}_{sol} = \mathbf{b}$ is said to be underdetermined. There are infinitely many solutions. It is the case where the number of observation is not large enough to solve the problem. But we can derive a feasible solution set. The matrix $\mathbf{\Sigma}$ has this particular shape :

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{pmatrix}_{m \times n}$$

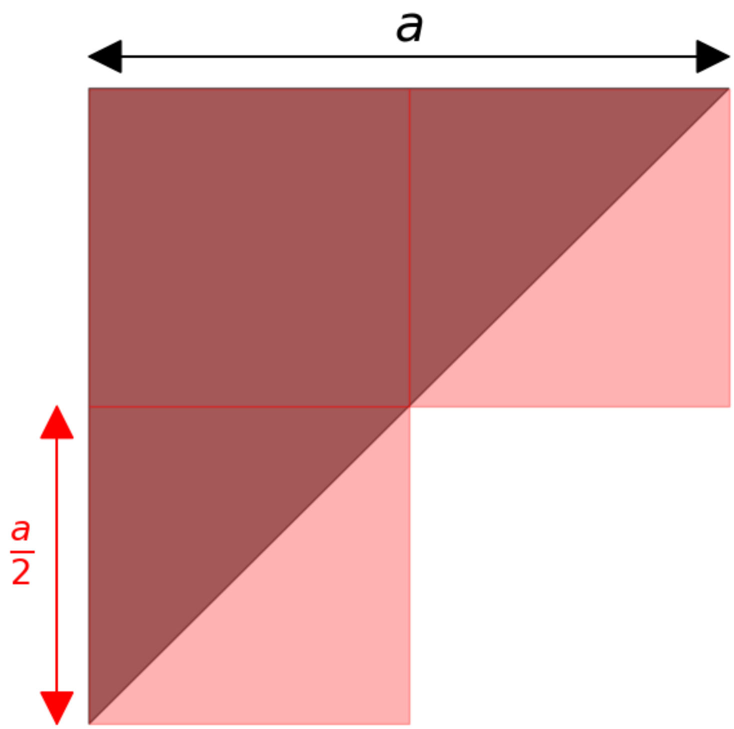

Figure 1: Squares of side $\frac{a}{2}$ realises a cover of the isosceles right triangle

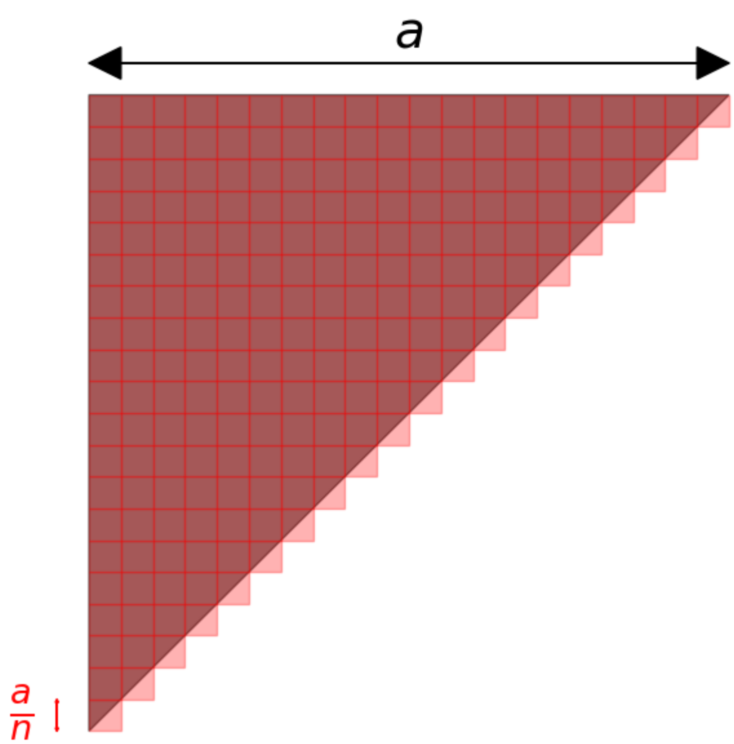

Figure 2: Stacking smaller squares gives a better covering of the isosceles right triangle

where $\forall\, i = 1, \dots, m, \quad \sigma_i > 0$.

To find a particular solution, we compute the squared error which is a function of $\mathbf{x} \in \mathbb{R}^m$:

$$\begin{aligned} e(\mathbf{x}) &= \|\mathbf{Ax} - \mathbf{b}\|^2 \\ &= (\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{b}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b} \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - 2\mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b} \end{aligned}$$

We compute the gradient and set it to zero to get the minimum of the squared error:

$$\nabla e(\mathbf{x_{part}}) = 2\mathbf{A}^T\mathbf{A}\mathbf{x_{part}} - 2\mathbf{A}^T\mathbf{b} = \mathbf{0}_{\mathbb{R}^n}$$

The issue compared to usual machine learning problem is that we cannot invert $\mathbf{A^TA}$ to derive the normal equation as rank$(\mathbf{A}) < n$. The SVD decomposition will help us to find an explicit particular solution. Let's define $\boldsymbol{\alpha} = \mathbf{V}^T\mathbf{x_{part}}$:

$$\begin{aligned} \mathbf{A}^T\mathbf{A}\mathbf{x_{part}} &= \mathbf{A}^T\mathbf{b} \\ (\mathbf{U\Sigma V}^T)^T\mathbf{U\Sigma V}^T\mathbf{x_{part}} &= (\mathbf{U\Sigma V}^T)^T\mathbf{b} \\ \mathbf{V\Sigma}^T\mathbf{U}^T\mathbf{U\Sigma V}^T\mathbf{x_{part}} &= \mathbf{V\Sigma}^T\mathbf{U}^T\mathbf{b} \\ \mathbf{\Sigma}^T\mathbf{\Sigma}\boldsymbol{\alpha} &= \mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} \\ \forall\, i = 1, \dots, m, \quad \sigma_i^2\alpha_i &= \sigma_i\mathbf{u}_i^T\mathbf{b} \\ \forall\, i = 1, \dots, m, \quad \alpha_i &= \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i} \end{aligned}$$

We derive $\mathbf{x}$ from the α_i, by taking $\alpha_i = 0$ if $i = m+1, \dots, n$:

$$\mathbf{x_{part}} = \mathbf{V}\boldsymbol{\alpha} = \sum_{i=1}^{m} \alpha_i\mathbf{v}_i = \sum_{i=1}^{m} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i = \sum_{i=1}^{m} \mathbf{v}_i\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i} = \left(\sum_{i=1}^{m} \frac{1}{\sigma_i}\mathbf{v}_i\mathbf{u}_i^T\right)\mathbf{b}$$

The matrix in brackets is named the pseudoinverse, and it is noted $\mathbf{A}^\dagger$. We recognise the rank 1 matrix decomposition with :

$$\begin{aligned} \mathbf{A}^\dagger &= \sum_{i=1}^{m} \frac{1}{\sigma_i}\mathbf{v}_i\mathbf{u}_i^T \\ &= [\frac{1}{\sigma_1}\mathbf{v}_1|\dots|\frac{1}{\sigma_m}\mathbf{v}_m] * [\mathbf{u}_1|\dots|\mathbf{u}_m]^T \\ &= \mathbf{V\Sigma^\dagger U}^T \end{aligned}$$

where $\mathbf{\Sigma}^\dagger = \text{diag}(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_m}) \in \mathbb{R}^{n\times m}$. It can also be computed by the formula:

$$\begin{aligned} \mathbf{A}^T(\mathbf{AA}^T)^{-1} &= \mathbf{V\Sigma}^T\mathbf{U}^T(\mathbf{U\Sigma V}^T\mathbf{V\Sigma}^T\mathbf{U}^T)^{-1} \\ &= \mathbf{V\Sigma}^T\mathbf{U}^T(\mathbf{U}\text{diag}(\sigma_1^2, \dots, \sigma_p^2)\mathbf{U}^T)^{-1} \\ &= \mathbf{V\Sigma}^T\mathbf{U}^T\mathbf{U}\text{diag}(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_p^2})\mathbf{U}^T \\ &= \mathbf{V\Sigma^\dagger U}^T \\ &= \mathbf{A}^\dagger \end{aligned}$$

Now let's compute the general solution of the equation, so $\mathbf{Ax}_{gene} = \mathbf{0}_{\mathbb{R}^m}$. It is equivalent to derive the null space of $\mathbf{A}$ and, as $\mathbf{U}$ is invertible, equivalent to derive the null space of $\mathbf{\Sigma V}^T$. Let's define $\mathbf{x'} = \mathbf{V}^T\mathbf{x_{gene}}$. As $\mathbf{\Sigma}$ is diagonal in $\mathbb{R}^{m\times n}$, a vector in the null space has the shape $(0, \dots, 0, x'_{m+1}, \dots, x'_n)^T$. As $\mathbf{V}^T$ is orthogonal, the null space of $\mathbf{A}$ is span$(\mathbf{v}_{m+1}, \dots, \mathbf{v}_n)$.

We can also derive a formula based on $\mathbf{A}$. Let's define:

$$\mathbf{P} = \mathbf{I}_n - \mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_n - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$$

By substituting the respective SVD:

$$\begin{aligned}
\mathbf{P} &= \mathbf{I}_n - \mathbf{V}\mathbf{\Sigma}^\dagger \mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{I}_n - [\mathbf{v}_1|\dots|\mathbf{v}_n] \begin{pmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} [\mathbf{v}_1|\dots|\mathbf{v}_n]^T \\
&= \mathbf{I}_n - \sum_{i=1}^{m} \mathbf{v}_i\mathbf{v}_i^T
\end{aligned}$$

We decompose any vector $\mathbf{x}$ on the orthogonal basis formed by the $\mathbf{v}_i$: $\mathbf{x} - \sum_{j=1}^{n} \nu_j \mathbf{v}_j$. We can compute the image of $\mathbf{P}$:

$$\begin{aligned}
\mathbf{P}\mathbf{x} &= \sum_{j=1}^{n} \nu_j \mathbf{v}_j - \sum_{i=1}^{m} \left(\mathbf{v}_i\mathbf{v}_i^T \sum_{j=1}^{n} \nu_j \mathbf{v}_j \right) \\
&= \sum_{j=1}^{n} \nu_j \mathbf{v}_j - \sum_{i=1}^{m} \nu_i \mathbf{v}_i \\
&= \sum_{j=m+1}^{n} \nu_j \mathbf{v}_j
\end{aligned}$$

We have found that the image of $\mathbf{P}$ is the null space of $\mathbf{A}$.

Finally $\mathbf{x}_{sol} = \mathbf{x}_{part} + \mathbf{x}_{gene}$:

$$\boxed{\mathbf{x}_{sol} = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^T\mathbf{b} + \operatorname{span}(\mathbf{v}_{m+1}, \dots, \mathbf{v}_n)} \tag{1}$$

$$\boxed{\mathbf{x}_{sol} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} + (\mathbf{I}_n - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\mathbf{x} \quad \forall \mathbf{x}} \tag{2}$$

3 Question 3

To compute the gradient of $E(u) = \sum_{i=1}^{2}\sum_{j=1}^{2} \arctan(u_{i,j} - u_{i-1,j})$ from sums starting from **1** with the assumption that $u_{0,1} = u_{0,2} = 0$, we need to compute the partial derivative with respect to $u_{1,1}$, $u_{1,2}$, $u_{2,1}$ and $u_{2,2}$.

For $u_{1,1}$:

$$\begin{aligned}
\frac{\partial E}{\partial u_{1,1}} &= \frac{\partial}{\partial u_{1,1}} \left[\arctan(u_{1,1} - 0) + \arctan(u_{2,1} - u_{1,1})\right] \\
&= \frac{1}{1 + u_{1,1}^2} + \frac{-1}{1 + (u_{2,1} - u_{1,1})^2}
\end{aligned}$$

For $u_{1,2}$:

$$\begin{aligned}
\frac{\partial E}{\partial u_{1,2}} &= \frac{\partial}{\partial u_{1,2}} \left[\arctan(u_{1,2} - 0) + \arctan(u_{2,2} - u_{1,2})\right] \\
&= \frac{1}{1 + u_{1,2}^2} + \frac{-1}{1 + (u_{2,2} - u_{1,2})^2}
\end{aligned}$$

For $u_{2,1}$:

$$\begin{aligned}\frac{\partial E}{\partial u_{2,1}} &= \frac{\partial}{\partial u_{2,1}}\left[\arctan(u_{2,1}-u_{1,1})\right]\\ &= \frac{1}{1+(u_{2,1}-u_{1,1})^2}\end{aligned}$$

For $u_{2,2}$:

$$\begin{aligned}\frac{\partial E}{\partial u_{2,2}} &= \frac{\partial}{\partial u_{2,2}}\left[\arctan(u_{2,2}-u_{1,2})\right]\\ &= \frac{1}{1+(u_{2,2}-u_{1,2})^2}\end{aligned}$$

Finally :

$$\nabla E(u) = \begin{pmatrix} \frac{\partial E}{\partial u_{1,1}} & \frac{\partial E}{\partial u_{1,2}} \\ \frac{\partial E}{\partial u_{2,1}} & \frac{\partial E}{\partial u_{2,2}} \end{pmatrix}$$

$$\boxed{\nabla E(u) = \begin{pmatrix} \frac{1}{1+u_{1,1}^2} + \frac{-1}{1+(u_{2,1}-u_{1,1})^2} & \frac{1}{1+u_{1,2}^2} + \frac{-1}{1+(u_{2,2}-u_{1,2})^2} \\ \frac{1}{1+(u_{2,1}-u_{1,1})^2} & \frac{1}{1+(u_{2,2}-u_{1,2})^2} \end{pmatrix}}$$

4 Question 4

Given the probability density function :

$$p(x;\alpha,\epsilon) = \frac{1}{Z}e^{-\alpha x^2-\epsilon}$$

Knowing that the $x(i)$ are independent and identically distributed samples, the likelihood function $\mathcal{L}$ is :

$$\begin{aligned}\mathcal{L}(\alpha; x(1),\ldots,x(m),\epsilon) &= \prod_{i=1}^{m} p(x(i);\alpha,\epsilon).\\ &= \frac{1}{Z^m}e^{-\sum_{i=1}^{m}(\alpha x(i)^2)-m\epsilon}\end{aligned}$$

The likelihood function is maximum when α is as small as possible :

$$\boxed{\hat{\alpha}_{MLE} \to 0}$$

It is clearly not a good estimator as it does not depend on $x(i)$.

We could try the first moment method, but :

$$E[X] = \int_{-\infty}^{\infty} x\frac{1}{Z}e^{-\alpha x^2-\epsilon}\,dx = 0$$

as the integrand is an odd function.

For the second moment :

$$E[X^2] = \int_{-\infty}^{\infty} x^2\frac{1}{Z}e^{-\alpha x^2-\epsilon}\,dx$$

Let's do an integration by parts:

$$u = \frac{x}{Z} \quad u' = \frac{1}{Z}$$
$$v = \frac{1}{-2\alpha}e^{-\alpha x^2 - \epsilon} \quad v' = xe^{-\alpha x^2 - \epsilon}$$

We get the primitive part equals to zero and:

$$E[X^2] = \frac{Z}{Z2\alpha} = \frac{1}{2\alpha}$$

Applying the method of the moments:

$$\boxed{\hat{\alpha}_{MM} = \frac{1}{2}\frac{n}{\sum_{i=1}^{m} x(i)^2}}$$