D5: Databases and Information Systems
Data Mining and Matrices, SS 2017
Homework #3: Optimization
Tutorial: **7 June 2017** at 10:15

You can discuss these problems with other students, but everybody must do and present their own answers. You can use computers etc. to perform the algebraic operations, but you must show the intermediate steps (and "computer said so" is never a valid answer). You are of course free to use material from the Internet, but again, you must present the intermediate steps and you must also be able to explain why the steps are valid and why you chose them. You can mark an answer even if it is not complete or correct, as long as you have made significant progress towards solving it. Note, however, that the TA does the final decision on whether your solution is complete (or correct) enough for a mark.

Problem 1 (More on pseudo-inverse). Let $\boldsymbol{A} \in \mathbb{R}^{n\times k}$, where $n > k$. Remember that the *pseudo-inverse* of $\boldsymbol{A}$, $\boldsymbol{A}^+$, is defined as $\boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$, where $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is the SVD of $\boldsymbol{A}$ and $(\boldsymbol{\Sigma}^+)_{ii} = \frac{1}{\sigma_i}$ if $\sigma_i > 0$ and 0 otherwise. Show that $\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ when $\boldsymbol{A}^T\boldsymbol{A}$ is invertible.

Problem 2 (Hessians and Jacobians). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function with a well-defined Hessian (that is, all partial derivatives $\frac{\partial^2}{\partial x_i \partial x_j} f(\boldsymbol{x})$ exist). Show that $\boldsymbol{H}(f(\boldsymbol{x})) = \boldsymbol{J}(\nabla f(\boldsymbol{x}))^T$.

Problem 3 (Jacobian of an affine map). The chain rule for the gradient says that if $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}^n$, then

$$\nabla(f \circ g)(\boldsymbol{x}) = \boldsymbol{J}(g(\boldsymbol{x}))^T(\nabla f(\boldsymbol{y})) , \tag{3.1}$$

where $\boldsymbol{y}$ is the value of g at $\boldsymbol{x}$, $\boldsymbol{y} = g(\boldsymbol{x})$. This was used in the lecture to calculate the gradient of $\frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\right\|_F^2$. Calculate the Jacobian of the affine map $g(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$.

Problem 4 (Eigendecompositions and positive semidefiniteness). An *eigendecomposition* of a symmetric square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has the form $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix that has the *eigenvectors* of $\boldsymbol{A}$ in its columns, and $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ is a diagonal matrix that has the *eigenvalues* of $\boldsymbol{A}$ in its diagonal. Eigendecomposition, much like SVD, is unique: if $\boldsymbol{A}$ is symmetric real-valued matrix, and $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ for orthogonal $\boldsymbol{Q}$ and diagonal $\boldsymbol{\Lambda}$, then $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ is the eigendecomposition of $\boldsymbol{A}$.

In the lecture, it was said that a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is *positive semidefinite* if $\boldsymbol{z}^T\boldsymbol{A}\boldsymbol{z} \geq 0$ for all $\boldsymbol{z} \in \mathbb{R}^n$. An equivalent definition for symmetric matrices is that all of its eigenvalues are nonnegative. Let $\boldsymbol{A} \in \mathbb{R}^{n\times k}$ be an arbitrary real-valued matrix. Show that matrices $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{C} = \boldsymbol{A}^T\boldsymbol{A}$ are positive semidefinite. (*Hint:* One way to do this is to use SVD and the uniqueness of the eigendecomposition.)

Problem 5 (Convexity of least-squares regression). Let $\boldsymbol{A} \in \mathbb{R}^{n\times k}$ and $\boldsymbol{b} \in \mathbb{R}^n$ be given. Show that the problem of finding $\boldsymbol{x} \in \mathbb{R}^k$ that minimizes $\frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\right\|_2^2$ is *convex*. You can use the semidefiniteness results of Problem 4 even if you didn't solve the problem.

Problem 6 (More merry gradients). In most real-world matrix factorization applications, we want to *regularize* the factor matrices. For example, we might want to find matrices that have a low Frobenius norm. In these cases, our loss function takes the form

$$L(\boldsymbol{B}, \boldsymbol{C}; \boldsymbol{A}, \lambda_1, \lambda_2) = \frac{1}{2}\left\|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{C}\right\|_F^2 + \lambda_1\left\|\boldsymbol{B}\right\|_F^2 + \lambda_2\left\|\boldsymbol{C}\right\|_F^2 , \tag{6.1}$$

where $\lambda_i \in \mathbb{R}$ are the *regularization coefficients*. Show that the loss function L is convex when either $\boldsymbol{B}$ or $\boldsymbol{C}$ is fixed, and calculate the corresponding gradients.

max planck institut
informatik