MLISP: Machine Learning in Signal Processing

Solutions to problem set 4

Prof. V. I. Morgenshtern

Solver: S. Subic, S. Lotter, V. I. Morgenshtern

Problem 1: Reduction of convex optimization problems to standard form

First, transform the objective:

$$\begin{aligned}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 &= (\mathbf{A}\mathbf{x}-\mathbf{b})^{\mathsf{T}}(\mathbf{A}\mathbf{x}-\mathbf{b})\\ &= (\mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}-\mathbf{b}^{\mathsf{T}})(\mathbf{A}\mathbf{x}-\mathbf{b})\\ &= \mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathbf{x}-\mathbf{b}^{\mathsf{T}}\mathbf{A}\mathbf{x}-\mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{b}+\mathbf{b}^{\mathsf{T}}\mathbf{b}\\ &= \mathbf{x}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathbf{x}-2\mathbf{b}^{\mathsf{T}}\mathbf{A}\mathbf{x}+\mathbf{b}^{\mathsf{T}}\mathbf{b}.\end{aligned}$$

Set

$$\mathbf{P}=2\mathbf{A}^{\mathsf{T}},\quad \mathbf{c}^{\mathsf{T}}=2\mathbf{b}^{\mathsf{T}}\mathbf{A},\quad d=\mathbf{b}^{\mathsf{T}}\mathbf{b}$$

and observe that the objective is transformed to standard form.

Second, transform the constraints. Set

$$\mathbf{G}=\begin{bmatrix}\mathbf{I}_{nxn} & \mathbf{0}_{nxn}\\ \mathbf{0}_{nxn} & -\mathbf{I}_{nxn}\end{bmatrix},\quad \mathbf{h}=\begin{bmatrix}\mathbf{u}\\ -\mathbf{l}\end{bmatrix}_{2nx1}$$

and observe that $\mathbf{l}\preceq\mathbf{x}\preceq\mathbf{u}$ is equivalent to $\mathbf{G}\mathbf{x}\preceq\mathbf{h}$.

Problem 2: Chernoff bound

1. We have to prove that:

$$e^{sX}\geq e^{sv}I_v(X),\ \text{for all } s\geq 0. \tag{1}$$

   Let's first look at the case when $X\geq v$. In this case, $I_v(X)=1$, and therefore (1) is equivalent to :

$$e^{sX}\geq e^{sv}$$

   which is true since $sX\geq sv$ because $s\geq 0$.

   In the case when $X<v$, we have $I_v(X)=0$ so that (1) is equivalent to

$$e^{sX}\geq 0$$

   which is always true.

2. From the previous point we know that

$$e^{sX}\geq e^{sv}I_v(X),\ \text{for all } s\geq 0.$$

taking the expectation on both sides of this inequality we have

$$\mathbb{E}[e^{sX}] \geq e^{sv}\mathbb{E}[I_v(X)], \text{ for all } s \geq 0.$$

Clearly,

$$\mathbb{E}[I_v(X)] = \int_{-\infty}^{\infty} I_v(x)p_X(x)dx = \int_{X \geq v} I_v(x)p_X(x)dx = \mathbb{P}[X \geq v].$$

and therefore

$$\mathbb{E}[e^{sX}] \geq e^{sv}\mathbb{P}[X \geq v].$$

Since the inequality holds for all $s \geq 0$ we conclude:

$$\mathbb{P}[X > v] \leq \min_{s \geq 0} e^{-sv}\mathbb{E}[e^{sX}]$$

.

3. Now we want to show that for a Gaussian variable X (with zero mean and unit variance):

$$\mathbb{P}[X > v] \leq e^{-\frac{v^2}{2}}.$$

Let's compute the the characteristic function of X:

$$\mathbb{E}[e^{sX}] = e^{\frac{s^2}{2}}.$$

Combining the result from the point 2 with this, we get:

$$\mathbb{P}[X > v] \leq e^{-sv}\mathbb{E}[e^{sX}] = e^{-sv}e^{\frac{s^2}{2}}.$$

Now in order to get the tightest bound, we minimize the right-hand side over s. To find the minimum, compute the derivative of the right-hand side and set it to zero:

$$-ve^{-sv}e^{\frac{s^2}{2}} + se^{-sv}e^{\frac{s^2}{2}} = 0 \implies (s-v)e^{-sv}e^{\frac{s^2}{2}} = 0 \implies s = v.$$

Therefore, the tightest possibe bound is:

$$\mathbb{P}[X > v] \leq e^{-v^2}e^{\frac{v^2}{2}} = e^{\frac{v^2}{2}}.$$

4. Following the same steps as in the previous point we have:

$$\begin{aligned}
\mathbb{P}[k - Q \geq t] &\leq \min_{s \geq 0} e^{-st} \int_{-\infty}^{k} \frac{1}{2^{k/2}\Gamma(k/2)}(k-q)^{k/2-1}e^{-\frac{k-1}{2}}e^{sq}dq \\
&= \min_{s \geq 0} e^{-st} \int_{0}^{\infty} \frac{1}{2^{k/2}\Gamma(k/2)}e^{-q/2}e^{s(k-q)}q^{k/2-1}dq \\
&= \min_{s \geq 0} e^{s(k-t)} \int_{0}^{\infty} \frac{1}{2^{k/2}\Gamma(k/2)}e^{-(1+2s)q/2}(1+2s)^{k/2-1}(1+2s)^{-(k/2-1)}ds \\
&= \min_{s \geq 0}(1+2s)^{-k/2}e^{s(k-t)}.
\end{aligned}$$

Optimizing w.r.t. s yields

$$
\begin{aligned}
&-\frac{k}{2}(1+2s)^{-k/2-1}2e^{s(k-t)}+(1+2s)^{-k/2}(k-t)e^{s(k-t)}=0 \Leftrightarrow \\
&e^{s(k-t)}(1+2s)^{-k/2}\left(-\frac{k}{1+2s}+k-t\right)=0 \Leftrightarrow \\
&2s(k-t)=0 \Leftrightarrow \\
&s=\frac{t}{2(k-t)}.
\end{aligned}
$$

Therefore, for $0 < t < k$:

$$
\mathbb{P}[Q \le k-t] = \mathbb{P}[k-Q \ge t] \le \left(1+\frac{t}{k-t}\right)^{-k/2} e^{t/2}.
$$

Problem 3: Cost function of logistic regression is convex

We will separately show that the following functions are convex

$$
-\sum_{i=1}^{n} y^{(i)} \log h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}), \quad -\sum_{i=1}^{n}(1-y^{(i)})\log\left(1-h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})\right)
$$

and then use the fact that the sum of convex functions is convex.

First, let's prove that

$$
f_1(\boldsymbol{\theta}) = \sum_{i=1}^{n} -y^{(i)} \log h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})
$$

is a convex function of $\boldsymbol{\theta}$. To do so it is sufficient to establish that every term in the sum is a convex function of $\boldsymbol{\theta}$. To prove that

$$
f_2(\boldsymbol{\theta}) = -y \log h_{\boldsymbol{\theta}}(\mathbf{x})
$$

we will prove that the Hessian matrix of $f_2(\boldsymbol{\theta})$ is positive semidefinite:

$$
\frac{\partial(-y\log(h_{\boldsymbol{\theta}}(\mathbf{x}))}{\partial\theta_j} = \frac{-y}{h_{\boldsymbol{\theta}}(\mathbf{x})} h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x}))x_j = -y(1-h_{\boldsymbol{\theta}}(\mathbf{x}))x_j
$$

so that

$$
\frac{\partial^2(-y\log(h_{\boldsymbol{\theta}}(\mathbf{x})))}{\partial\theta_j\partial\theta_k} = y h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x}))x_j x_k
$$

and finally

$$
\nabla^2 f_2(\boldsymbol{\theta}) = y h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x})) \begin{bmatrix} {x_1}^2 & x_1x_2 & \dots & x_1x_n \\ x_2x_n & {x_2}^2 & \dots & x_2x_n \\ \dots & \dots & \dots & \dots \\ x_nx_1 & x_nx_n & \dots & {x_n}^2 \end{bmatrix} = y h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x}))\mathbf{x}\mathbf{x}^{\mathsf{T}}.
$$

This matrix is positive semidefinite, because $y h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x}))$ is a nonnegative scalar and $\mathbf{x}\mathbf{x}^{\mathsf{T}}$ is a positive semidefinite matrix.

To prove that

$$f_3(\boldsymbol{\theta}) = -\sum_{i=1}^{n}(1-y^{(i)})\log\left(1-h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})\right)$$

is convex we proceed in a similar way. Note that

$$\frac{\partial \log(h_{\boldsymbol{\theta}}(1-\mathbf{x}))}{\partial\theta_j} = -\frac{1}{1-h_{\boldsymbol{\theta}}(\mathbf{x})}h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x}))x_j = -h_{\boldsymbol{\theta}}(\mathbf{x})x_j$$

so that

$$\frac{\partial^2(-(1-y)\log(h_{\boldsymbol{\theta}}(1-\mathbf{x})))}{\partial\theta_j\partial\theta_k} = (1-y)h_{\boldsymbol{\theta}}(\mathbf{x})(1-h_{\boldsymbol{\theta}}(\mathbf{x}))x_jx_k$$

from which the convexity follows in exactly the same way as above.

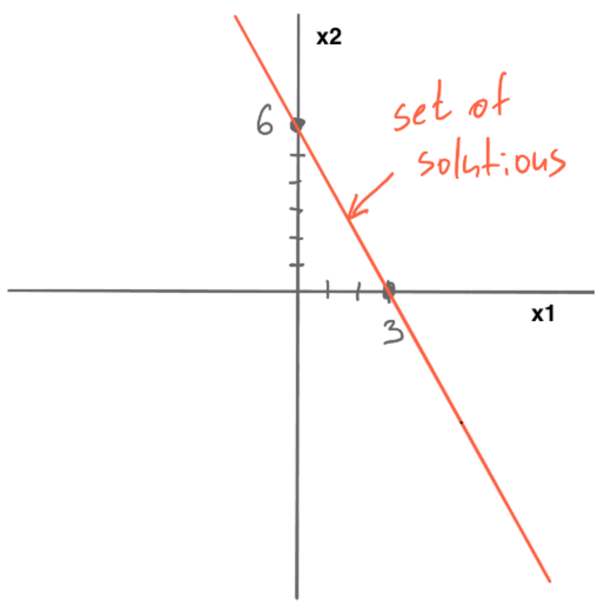
Problem 8: Solving underdetermined systems of equations (exam practice)

1. The equation can be written as

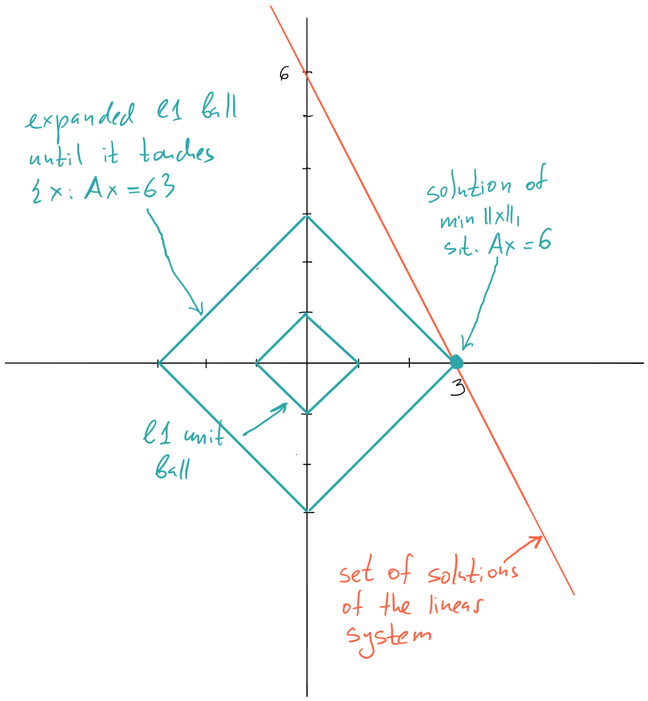
   $$2x_1 + x_2 = 6.$$

   Therefore the set of solutions is $\{[x_1\ x_2] : x_2 = 6 - 2x_1\}$.

   Here is this set:



2. The solution is the point $\mathbf{x} = [3\ 0]$. The following picture contains the graphical proof:

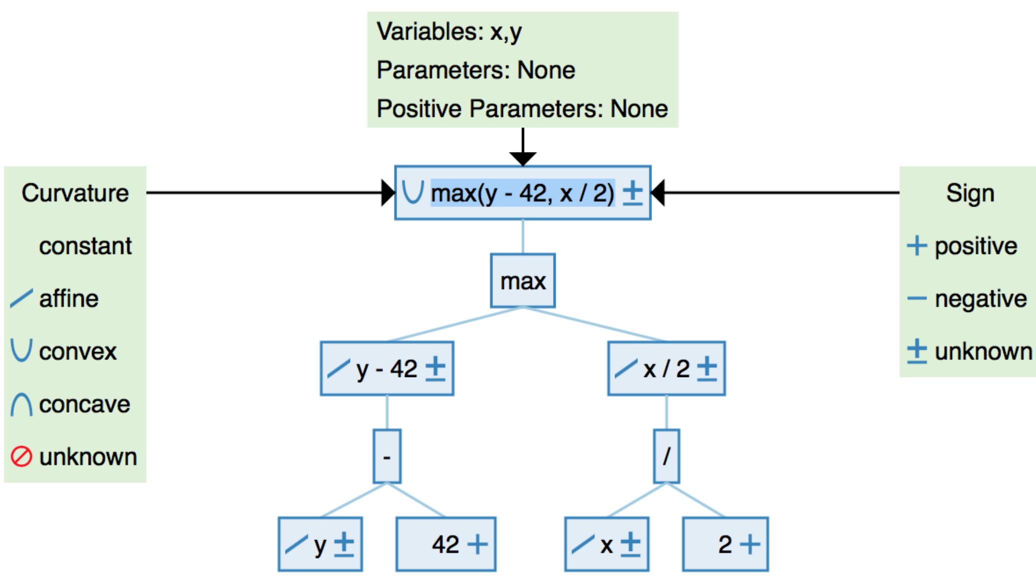


In the figure, the unit l1 ball is depicted. This ball is expanded in a self-similar way until it first hits the set $\{[x_1\ x_2] : x_2 = 6 - 2x_1\}$.

3. The solution vector $\mathbf{x} = [3\ 0]$ is one-sparse, because it contains only one nonzero element.

Problem 9: Convex optimization problems (exam practice)

1. The function is the composition of basic functions. It is convex, as follows from the disciplined convex programming decomposition diagram:



2. Here we optimize a convex (quadratic) function subject to linear (and hence convex) constraints. Therefore, this is a convex optimization problem.

3. Here we optimize a convex (linear) function subject to nonconvex constraints. Therefore, this is not a convex optimization problems.

4. Here we optimize a convex (linear) function subject to convex constraints. Therefore, this is a convex optimization problems.