Exercises
Computational Intelligence Lab
SS 2019

Machine Learning Institute
Dept. of Computer Science, ETH Zürich
Prof. Dr. Thomas Hofmann
Web http://cil.inf.ethz.ch/

Series 11, May 16-17, 2018
(Dictionary Learning and Compressed Sensing)

Problem 1 (Sparse coding with overcomplete dictionary):

Given a signal $\mathbf{x} = (3, 1, -2) \in \mathbb{R}^3$ and an overcomplete dictionary $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4] \in \mathbb{R}^{3\times4}$,

$$\mathbf{U} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix},$$

find the sparse representation $\mathbf{z}$ of the signal $\mathbf{x}$ with $\|\mathbf{z}\|_0 \leq 2$.

a. Find the atom $\mathbf{u}^{(1)}$ that minimize the reconstruction error $\|\mathbf{x} - \hat{\mathbf{x}}^{(0)}\|_2^2$ where $\hat{\mathbf{x}}^{(0)} = z^{(1)}\mathbf{u}^{(1)}$, and compute the residual $\mathbf{r}^{(1)} = \mathbf{x} - \hat{\mathbf{x}}^{(0)}$.
Solution: The atom $\mathbf{u}^{(1)}$ that minimizes the reconstruction error $\|\mathbf{x} - z^{(1)}\mathbf{u}^{(1)}\|$ is the atom that is best correlated $\mathbf{x}$. The correlation between the signal and the atoms in the dictionary are as following,

$$\begin{aligned} \langle \mathbf{x}, \mathbf{u}_1 \rangle &= \frac{2}{\sqrt{3}} \\ \langle \mathbf{x}, \mathbf{u}_2 \rangle &= -\frac{4}{\sqrt{3}} \\ \langle \mathbf{x}, \mathbf{u}_3 \rangle &= 0 \\ \langle \mathbf{x}, \mathbf{u}_4 \rangle &= \frac{6}{\sqrt{3}}. \end{aligned}$$

Since the absolute correlation coefficient between the atom $\mathbf{u}_4$ and the signal $\mathbf{x}$ has the largest value, $\hat{\mathbf{x}}^{(0)} = \langle \mathbf{x}, \mathbf{u}_4 \rangle \cdot \mathbf{u}_4 = \frac{6}{\sqrt{3}} \cdot \mathbf{u}_4$ minimizes $\|\mathbf{x} - \hat{\mathbf{x}}^{(0)}\|_2^2$. The residual becomes $\mathbf{r}^{(1)} = \mathbf{x} - \hat{\mathbf{x}}^{(0)} = (1, -1, 0)$.

b. Find the atom $\mathbf{u}^{(2)}$ that minimize the reconstruction error $\|\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(1)}\|_2^2$ where $\hat{\mathbf{x}}^{(1)} = z^{(2)}\mathbf{u}^{(2)}$.
Solution: Similarly, we want to find the atom best correlated with $\mathbf{r}^{(1)}$ among the remaining atoms $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ (why can't it be $\mathbf{u}_4$?). The correlation coefficients between the atoms and the residual are

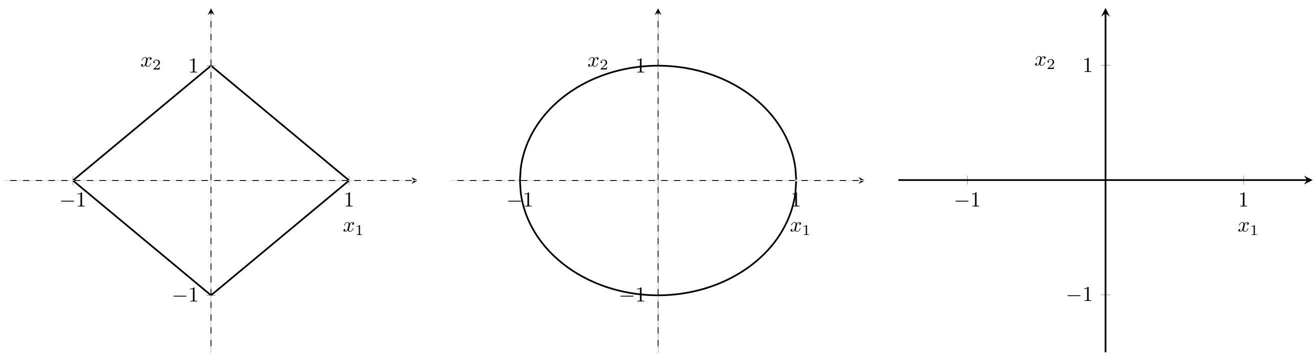
$$\begin{aligned} \langle \mathbf{r}^{(1)}, \mathbf{u}_1 \rangle &= 0 \\ \langle \mathbf{r}^{(1)}, \mathbf{u}_2 \rangle &= -\frac{2}{\sqrt{3}} \\ \langle \mathbf{r}^{(1)}, \mathbf{u}_3 \rangle &= \frac{2}{\sqrt{3}}. \end{aligned}$$

Because the absolute correlation coefficient values of $\mathbf{u}_2$ and $\mathbf{u}_3$ are the same, either $\hat{\mathbf{x}}^{(1)} = -\frac{2}{\sqrt{3}}\mathbf{u}_2$ or $\hat{\mathbf{x}}^{(1)} = \frac{2}{\sqrt{3}}\mathbf{u}_3$ can minimize $\|\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(1)}\|_2^2$.

c. Write down the sparse representation $\mathbf{z}$ of signal $\mathbf{x}$.
Solution: The sparse representations $\mathbf{z}$ that satisfy $\|\mathbf{z}\|_0 \leq 2$ are $(0, 0, 0, \frac{6}{\sqrt{3}})$, $(0, 0, \frac{2}{\sqrt{3}}, \frac{6}{\sqrt{3}})$ and $(0, -\frac{2}{\sqrt{3}}, 0, \frac{6}{\sqrt{3}})$.

Problem 2 (Compressed Sensing):

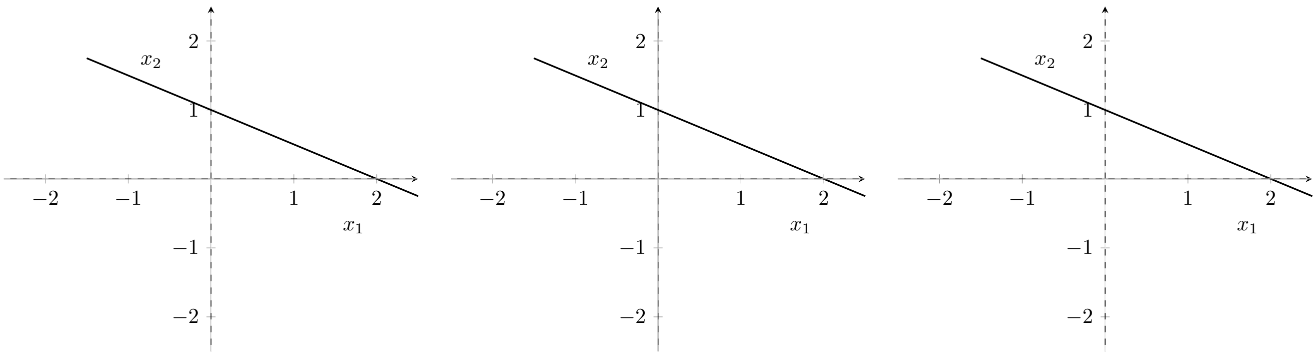
a. Map each of the three equations $\|\mathbf{x}\|_2 = 1$, $\|\mathbf{x}\|_1 = 1$, and $\|\mathbf{x}\|_0 = 1$ to a plot among a., b., or c. on the following figure. Note that $\mathbf{x}$ is s 2D vector with coordinates x_1 and x_2 (i.e. $\mathbf{x} = \begin{bmatrix} x_1, x_2 \end{bmatrix}$).



a. $\|\mathbf{x}\|_1$ b. $\|\mathbf{x}\|_2$ c. $\|\mathbf{x}\|_0$

b. Show the solution of each optimization problem on plots a., b., and c. of the following figure.

$\min \|\mathbf{x}\|_2$ Subject to $\frac{1}{2}x_1 + x_2 = 1$	$\min \|\mathbf{x}\|_1$ Subject to $\frac{1}{2}x_1 + x_2 = 1$	$\min \|\mathbf{x}\|_0$ Subject to $\frac{1}{2}x_1 + x_2 = 1$



a. b. c.

Solution:

a. The answer will be the closest point on the line to the origin, i.e.

$$\frac{1}{2}\mathbf{x}_1 + \mathbf{x}_2 = 1 \leftrightarrow \mathbf{x}_1 = 2 - 2\mathbf{x}_2 \tag{1}$$

$$\min\{\mathbf{x}_1^2 + \mathbf{x}_2^2\} = \min\{(2 - 2\mathbf{x}_2)^2 + \mathbf{x}_2^2\} \tag{2}$$

$$\frac{d}{d\mathbf{x}_2}\left[(2 - 2\mathbf{x}_2)^2 + \mathbf{x}_2^2\right] \stackrel{!}{=} 0 \leftrightarrow \mathbf{x}_2 = 0.8, \mathbf{x}_1 = 0.4 \tag{3}$$

b. $\mathbf{x}_1 = 0, \mathbf{x}_2 = 1$ c. two solutions $[\mathbf{x}_1 = 0, \mathbf{x}_2 = 1], [\mathbf{x}_1 = 2, \mathbf{x}_2 = 0]$

c. We can formulate the above three optimization problem as

$$\min \|\mathbf{x}\|_p$$
$$\text{subject to } \frac{1}{2}x_1 + x_2 = 1,$$

where $p \in \{0, 1, 2\}$. Mark the right sentence using your previous answers.

[ ] Solutions of the constrained problems have intersection for $p = 1$ and $p = 0$.

[ ] Solutions of the constrained problems have intersection for $p = 2$ and $p = 0$.

Solution:

Solutions of the constrained problems have intersection for $p = 1$ and $p = 0$.

Problem 3 (Compressed Sensing):

Please find the iPython notebook Compressed_sensing.ipynb from answer the question in this file.

Solution: Please find the solution in the same directory.

Problem 4 (Matching pursuit algorithm):

a. Find an overcomplete dictionary and a vector $\mathbf{x}$ such that the approximation $\hat{\mathbf{x}}$ resulting from the matching pursuit algorithm will never exactly equal $\mathbf{x}$ no matter the number of iterations. Note that this implies that at least one atom will be selected multiple times.

Solution : Consider the dictionary consisting of atoms $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (\sqrt{2}/2, \sqrt{2}/2)$ and $\mathbf{u}_3 = (\sqrt{3}/2, 1/2)$. For vector $(0, 1)$, we argue that the residual is always in one of the following two forms:

- $\mathbf{v}(x) = (0, x)$ for some $x \neq 0$
- or $\mathbf{w}(x) = (-x, x)$ for some $x \neq 0$.

For the first form, atom $\mathbf{u}_2$ always gets selected as

$$|\langle \mathbf{v}(x), \mathbf{u}_2 \rangle| > |\langle \mathbf{v}(x), \mathbf{u}_3 \rangle| > |\langle \mathbf{v}(x), \mathbf{u}_1 \rangle|.$$

Thus, the new residual is

$$\mathbf{v}(x) - \langle \mathbf{v}(x), \mathbf{u}_2 \rangle \mathbf{u}_2 = \mathbf{v}(x) - x\frac{\sqrt{2}}{2}\mathbf{u}_2 = \mathbf{w}(x/2).$$

For the second form, atom $\mathbf{u}_1$ always gets selected as

$$|\langle \mathbf{w}(x), \mathbf{u}_1 \rangle| > |\langle \mathbf{w}(x), \mathbf{u}_3 \rangle| > |\langle \mathbf{w}(x), \mathbf{u}_2 \rangle|.$$

Thus, the new residual is

$$\mathbf{w}(x) - \langle \mathbf{w}(x), \mathbf{u}_1 \rangle \mathbf{u}_1 = \mathbf{w}(x) - (-x)\mathbf{u}_1 = \mathbf{v}(x).$$

As the initial residual $\mathbf{v}(1) = (0, 1)$ is of the first form, the residual energy will never exactly be 0 by induction.

Note that the atom $\mathbf{u}_3$ was never used here, the atom was only added to make the dictionary overcomplete.

b. Find an instance where the sparse representation returned by matching pursuit (assuming that after some number of iterations, the approximation is perfect) is not optimal, i.e. there is a different representation for which the 0-norm is strictly smaller.

Solution: Consider the dictionary consisting of atoms $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (0, 1)$ and $\mathbf{u}_3 = (\sqrt{2}/2, \sqrt{2}/2)$. Clearly, the sparsest representation of $\mathbf{x} = (2, 1)$ is $\mathbf{x} = 2\mathbf{u}_1 + \mathbf{u}_2 + 0\mathbf{u}_3$ or written in vector form $(2, 1, 0)$.

Let us now execute matching pursuit on $\mathbf{x}$. As $|\langle \mathbf{x}, \mathbf{u}_1 \rangle| = 2$, $|\langle \mathbf{x}, \mathbf{u}_2 \rangle| = 1$ and $|\langle \mathbf{x}, \mathbf{u}_3 \rangle| = \frac{3}{2}\sqrt{2}$, the vector $\mathbf{x}$ is initially projected on atom $\mathbf{u}_3$. The resulting residual becomes

$$\mathbf{r}_1 = \mathbf{x} - \langle \mathbf{x}, \mathbf{u}_3 \rangle \mathbf{u}_3 = (1/2, -1/2).$$

Evaluating $|\langle \mathbf{r}_1, \mathbf{u}_1 \rangle| = 1/2$, $|\langle \mathbf{r}_1, \mathbf{u}_2 \rangle| = 1/2$ and $|\langle \mathbf{r}_2, \mathbf{u}_3 \rangle| = 0$ shows that the residual $\mathbf{r}_1$ is projected on $\mathbf{u}_1$ (we could also project it on $\mathbf{u}_2$ here, the end result would be the same, just not the order of selected atoms). The residual therefore becomes

$$\mathbf{r}_2 = \mathbf{r}_1 - \langle \mathbf{r}_1, \mathbf{u}_1 \rangle \mathbf{u}_1 = (0, -1/2).$$

Evaluating $|\langle \mathbf{r}_2, \mathbf{u}_1 \rangle| = 0$, $|\langle \mathbf{r}_2, \mathbf{u}_2 \rangle| = 1/2$ and $|\langle \mathbf{r}_2, \mathbf{u}_3 \rangle| = 0$ shows that the residual $\mathbf{r}_2$ is projected on $\mathbf{u}_2$. The resulting residual becomes

$$\mathbf{r}_3 = \mathbf{r}_2 - \langle \mathbf{r}_2, \mathbf{u}_2 \rangle \mathbf{u}_2 = (0, 0).$$

As $\mathbf{r}_3 = \mathbf{0}$, the algorithm terminates. The sparse representation computed by matching pursuit is $(1/2, -1/2, \frac{3}{2}\sqrt{2})$ while the optimal sparse representation is $(2, 1, 0)$.