Introduction to Data Processing and Representation
(236201)
Spring 2024

Homework 1

- **Published date: 13/06/2024**
- **Deadline date: 29/06/2024 23:59**

Guidelines:

- Submission is in pairs only.
- Submit your entire solution (including the theoretical part, the Python part and the Python code) electronically via the course website. The file should be a zip file containing the your PDF submission and Python code.
- The submission should be in English and in a clear printed form (recommended) or a clear hand-writing.
- Rigorous mathematical proofs and reasoning are required for theoretical questions. Vague answers and unjustified claims will not be accepted.

In this exercise we revisit the L^p uniform sampling problem for p a real scalar $p \geq 1$. Let N bet the number of samples. We partition $[0,1]$ into N uniform intervals I_i. For a real function f defined on $[0,1]$, we sample it by considering piece-wise constant functions $\hat{f}$ constant on each I_i. The weighted L^p sampling problem consists in solving the following optimization problem:

$$\min_{\hat{f}} \mathcal{E}^p(f, \hat{f}) = \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^p w(x) dx,$$

subject to the sampling constraint: $\hat{f}$ needs to be constant on each interval I_i, and where $w(x) > 0$ is a strictly positive weight function independent of f and $\hat{f}$. In this exercise we assume we are provided with a method capable of computing integrals.

a. Assume here that w is a constant function. Give, without proof, what is the optimal $\hat{f}_p$ when $p = 1$ and when $p = 2$.

For P=2 the proble simplified for finding the peacewise functio $\hat{f}$ that minimiz SE

$$\min_{\hat{f}} \varepsilon^p(f, \hat{f}) = \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^p w(x) \, dx$$

W is constant therefor: =

$$E_p(f, \hat{f}) = w \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^p \, dx$$

for the case $\epsilon = 2$:

$$E_2(f, \hat{f}) = w \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^2 \, dx$$

for each interval I_i, $\hat{f}$ is constant, Denote this constant by X_i, now the objective function is

$$\int_{\frac{i-1}{N}}^{\frac{i}{N}} (f(x) - X_i)^2 dx$$

because we aiming to find the optimal respect to $\hat{f}$ let's calle the derivative respect to X_i's and set to zero

$$\frac{d}{dX_i}\left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} (f(x) - X_i)^2 \, dx \right) = -2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} (f(x) - X_i) \, dx = 0$$

$$\Rightarrow \int_{\frac{i-1}{N}}^{\frac{i}{N}} f(x) \, dx = Nc_i \rightarrow c_i = \frac{1}{N} \int_{\frac{i-1}{N}}^{\frac{i}{N}} f(x) \, dx$$

What is the MSE.

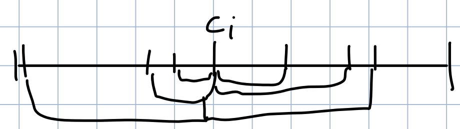
case $p=1$

$$\min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^1 dx = \underbrace{\sum_1^N \int_{\frac{i-1}{N}}^{\frac{i}{N}}}_{\hat{f} \text{ is constant}} |f(x) - c_i| \, dx$$

L1 (median)

for each integral should be minimize by collect the best representive c_i what is actually the median (we saw it at the lecture)

$$\int_{f(x)<c_i} (c_i - f(x))\,dx = \int_{f(x)>c_i} (f(x) - c_i)\,dx$$



So if M is odd $f\left(x_{(m+1)/2}\right) = m$

M is even $\dfrac{f(x_{m/2}) + f(x_{m/2+1})}{2} = m$

B). For general w, what is the optimal $\hat{f}_p$ when $p = 2$?

$$E_2(f, \hat{f}) = \int_0^1 |f(x) - \hat{f}(x)|^2 \, w(x)\,dx$$

Where $\hat{f}$ is constant for each I_i, sign $\hat{f}$ as c_i

$$\left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} (f(x) - c_i)^2 \, w(x)\,dx \right) \frac{d}{dc_i} = -2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} (f(x) - c_i)\, w(x) = 0$$

$$c_i = \frac{\int_{\frac{i-1}{N}}^{\frac{i}{N}} f(x) \cdot w(x)\,dx}{\int_{\frac{i-1}{N}}^{\frac{i}{N}} w(x)\,dx}$$

c. For general w, what is the optimal $\hat{f}_p$ when $p = 1$? You may use the same level of preciseness as in the lectures rather than in the tutorial.

c optimal $\hat{f}$ when $p=1$

for this case our objective is:

$$E_1(f, \hat{f}) = \int_0^1 |f(x) - \hat{f}(x)| \, w(x) \, dx$$

$$= \sum_{i=1}^{N} \int_{\frac{i-1}{N}}^{\frac{i}{N}} |f(x) - c_i| \, w(x_i) \, dx$$

so in this scenario optimal $\hat{f}$ on each I_i

will be the weighted values median of $f(x)$

on I, then the weighted median c_i is

a point such that

$$\int_{f(x)<c_i} w(x) \, dx = \int_{f(x)>c_i} w(x) \, dx$$

d. Prove that the optimization problem can be rewritten as a sum of N independent optimization problems depending solely on what happens in each interval. That is find $\mathcal{E}_i^p(f_i, \hat{f}_i)$ such that $\mathcal{E}^p(f, \hat{f}) = \sum_{i=1}^{N} \mathcal{E}_i^p(f_i, \hat{f}_i)$ where f_i and $\hat{f}_i$ are the functions f and $\hat{f}$ restrained to the interval I_i.

By given $E_p(f, \hat{f}) = \min_{f} \int_0^1 |f(x) - \hat{f}(x)|^p w(x)\, dx$

wher f is a real function defined on $[0,1)$, $\hat{f}$ piecewise constant funcion for each I_i and $w(x) > 0$ weight function, we can decomposit the integrals over the intervals $I_i = \left[\frac{i-1}{N}, \frac{1}{N}\right]$

Ther for

$$E_p(f, \hat{f}) = \sum_{i=1}^{N} \int_{\frac{i-1}{N}}^{\frac{i}{N}} |f(x) - c_i|^p w(x)\, dx$$

Then for each sub problem that defined on the intervall I_i the min proble will be:

$$E_{p_i}(f_i, \hat{f}_i) = \int_{\frac{i-1}{N}}^{\frac{i}{N}} |f(x) - c_i|^p w(x)\, dx$$

Whar the sum of the independent probleme is:

$$E_p(f, \hat{f}) = \sum_{i=1}^{N} E_{p_i}(f_i, \hat{f}_i)$$

e. As in the case where $p = 1$, explicitly computing the values of $\hat{f}_p$ is non-trivial when $p \neq 2$. We thus wish to use the simplicity of the L^2 optimization problem to solve the general L^p optimization. In this question, fix $i \in \{1, \cdots, N\}$ and work in I_i, thus focus on $\mathcal{E}_i^p$.

i. Assume that $f_i(x) \neq \hat{f}_i(x)$ for all $x \in I_i$. Find a positive function $w_{f_i,\hat{f}_i}$ depending on f_i and $\hat{f}_i$ such that $|f_i(x) - \hat{f}_i(x)|^p = w_{f_i,\hat{f}_i}(x)(f_i(x) - \hat{f}_i(x))^2$.

ii. Under the same assumption, rewrite the optimization of $\mathcal{E}_i^p$ as a weighted L^2-like optimization problem except that in this new formulation the positive weight function $w'_{f_i,\hat{f}_i}$ may depend on f_i and $\hat{f}_i$.

r

e) i) $$|f_i(x) - \hat{f}_i(x)|^p = w_{f_i,\hat{f}_j}(x)(f_i(x) - \hat{f}_i(x))^2$$

$$W_{f_i\hat{f}_j}(x) = \frac{|f_i(x) - \hat{f}_i(x)|^p}{\underbrace{|f_i(x) - \hat{f}_i(x)|^2}_{\neq 0}} = \boxed{|f_i(x) - \hat{f}_i(x)|^{p-2}}$$

ii)

According what we found such

$$w''_{f\cdot\hat{f}_i}(x) = |f_i(x) - \hat{f}_i(x)|^{p-2}$$

what convert the L_2-like problem into

$$E_{p,i}(f_i,\hat{f}_i) = \int_{I_i} W_{f_i,\hat{f}_i}(x)\,(f_i(x) - \hat{f}_i(x))^p$$

$$= E_{p,i}(f_i,\hat{f}_i) \underset{\downarrow\, w'}{=} \int_{I_i} w'_{f_i,\hat{f}_i}(x)\,\frac{(f_i(x) - \hat{f}_i(x))^p}{|f_i(x) - \hat{f}_i(x)|^{p-2}}\,dx$$

$$\Rightarrow E_{p,i}(f_i,\hat{f}_i) = \int_{I_i} W'_{f_i,\hat{f}_i}(x)\,|f_i(x) - \hat{f}_i(x)|^2\,dx$$

iii. Under the same assumption, solving this L^2-like optimization problem is hard because the $w'_{f_i,\hat{f}_i}$ is not necessarily independent of $\hat{f}_i$. It would be much simpler if the weight function was independent of it. Why?

When the wieght function w' depandes on $\hat{f}_i$ the L_2 problem become:

non-linearity: $\hat{f}_i$ effect on the weight altering

complex: each iteration requires recalculating the weights.

iv. When we remove the previous assumption, why do we prefer to use the function $\tilde{w}_{f_i,\hat{f}_i}(x) = \min\{\frac{1}{\varepsilon}, w_{f_i,\hat{f}_i}(x)\}$ instead of $w_{f_i,\hat{f}_i}(x)$, where $\varepsilon > 0$ a small fixed number?

To prevent overfiting and numerical instabilitY by capping excessivey large number wighes that occur when the error $|f_i(x) - \hat{f}_i(x)|$ very small for $p=100$

$$|f_i(x)-\hat{f}_i(x)|^{100-2} = \left(\frac{1}{|f_i(x)-\hat{f}_i(x)|}\right)^{100} \rightarrow$$ become to be vrary large

less sensitive than its diminish the influence of small discrepancis between $f,\hat{f}$

Algorithm: Iterative Stabilized Weight Optimization with Output

Jupiter notebook

Input:

- f_i: function defined on interval I_i
- `initial_guess`: initial guess for $\hat{f}_i$
- p: power used in the L_p norm
- `epsilon`: small positive number to cap the weights
- `max_iterations`: maximum number of iterations
- `tolerance`: small value for convergence check

Output:

- Final approximation $\hat{f}_{\text{next},i}$ for interval I_i after the last iteration

Procedure:

1. **Initialize:**
   - $\hat{f}_i = \text{initial\_guess}$
   - Calculate initial weights $w_i' = \min\left\{\frac{1}{\epsilon}, |f_i(x) - \hat{f}_i(x)|^{p-2}\right\}$
2. **Set iteration counter** $k = 0$
3. **Repeat:** a. Using fixed weights w_i', solve for $\hat{f}_{\text{next},i}$:
   - Compute the weighted average:

$$\hat{f}_{\text{next},i} = \left(\int_{I_i} w_i' \cdot f_i(x)\,dx\right) / \left(\int_{I_i} w_i'\,dx\right)$$

   b. **Check for convergence:**
   - If the change in $\hat{f}_{\text{next},i}$ from $\hat{f}_i$ is less than `tolerance` or $k \geq \text{max\_iterations}$:
     - **break**
   - Else:
     - Update $\hat{f}_i$ to $\hat{f}_{\text{next},i}$ c. Recalculate weights for the next iteration:
   - $w_i' = \min\left\{\frac{1}{\epsilon}, |f_i(x) - \hat{f}_i(x)|^{p-2}\right\}$ d. **Increment iteration counter** k
4. **Return the final approximation** $\hat{f}_{\text{next},i}$
   - Else:
     - Update $\hat{f}_i$ to $\hat{f}_{\text{next},i}$ c. Recalculate weights for the next iteration:
   - $w_i' = \min\left\{\frac{1}{\epsilon}, |f_i(x) - \hat{f}_i(x)|^{p-2}\right\}$ d. **Increment iteration counter** k
5. **Return the final approximation** $\hat{f}_{\text{next},i}$

f. Write a pseudo code for approximately solving the weighted L^p optimization problem using only L^2 optimizations.

g. What is the name of this algorithm? No points will be awarded to this question and we will not penalise the ignorant. IRLS

Algorithm: Iterative Reweighted Least Squares (IRLS) for Weighted L_p Optimization

Input:

- `f`: Function defined on [0, 1]
- `p`: The power used in the L_p norm, (p \geq 1)
- `epsilon`: A small positive number to stabilize weights
- `max_iterations`: Maximum number of iterations to prevent infinite loops
- `tolerance`: Convergence tolerance

Output:

- (\hat{f}): Approximation of (f) that minimizes the weighted L_p norm

Procedure:

1. **Initialize (\hat{f}) to a suitable initial guess (e.g., mean of (f))**
2. **Compute initial weights based on (\hat{f}):**
    - (w(x) = \min{1/\epsilon, |f(x) - \hat{f}(x)|^{p-2}})
3. **Set iteration counter (k = 0)**
4. **Repeat:** a. **Solve the weighted L_2 optimization problem to update (\hat{f}):**
    - (\hat{f} = \arg\min_{g} \int_0^1 w(x) (f(x) - g(x))^2 , dx) (This step often involves solving a system of linear equations or using a numerical optimizer)

    b. **Update weights (w(x)) using the new approximation (\hat{f}):**
    - (w(x) = \min{1/\epsilon, |f(x) - \hat{f}(x)|^{p-2}})

    c. **Check for convergence:**
    - If the change in (\hat{f}) is less than tolerance or (k) >= max_iterations:
        - **break**

    d. **Increment the iteration counter (k)**
5. **Return the optimized approximation (\hat{f})**

2. Signal Discretization using a Piecewise-Linear Approximation

In this problem, we extend the sampling procedure to rely on a piecewise-linear approximation of the signal. The given signal, $\phi(t)$, is defined for $t \in [0, 1)$ as a mapping to the range of values $[\phi_L, \phi_H]$.

Consider a discretization procedure based on a uniform segmentation of the unit interval into N intervals of equal size, i.e.,

$$\Delta_i = \left[\frac{i-1}{N}, \frac{i}{N}\right), \quad i = 1, ..., N \tag{1}$$

The approximated signal, $\hat{\phi}(t)$, is formed from linear approximations, each associated with an interval:

$$\hat{\phi}(t) = a_i(t - t_i) + c_i, \quad t \in \left[\frac{i-1}{N}, \frac{i}{N}\right), \tag{2}$$

where a_i and c_i are real-valued scalar constants defining the linear approximation of the i-th interval, and t_i is the center of the i-th interval.

The approximations are evaluated here for the MSE criterion.

a. Show that for a positive integer k:

$$\int_{t \in \Delta_i} (t - t_i)^k dt = \begin{cases} 0, & \text{if } k \text{ is odd} \\ \frac{|\Delta_i|^{k+1}}{2^k (k+1)}, & \text{if } k \text{ is even} \end{cases} \tag{3}$$

, where $|\Delta_i|$ is the size of the interval.

By given $\Delta_i = \left[\overset{e_1}{\frac{i-1}{N}}, \overset{e_0}{\frac{i}{N}}\right]$, t_i is the center of Δ_i

$t_i = \frac{2i-1}{2N}$, $|\Delta_i|$ the size of the Δ_i

Then the intervals is

$$\int_{e_0}^{e_1} (t - t_i)^K dt$$

For odd $\int_{0}^{0} f(t)\, dt = 0$. for k = odd

is negative symetric function then for symetric

$[a,b]$, $a = \frac{i-1}{N}$, $b = \frac{i}{N}$ then

$$\int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^k \, dt = 0$$

k = even, let define $u = t - t_i$, $du = dt$

then $t = \frac{i-1}{N} \to \frac{i}{N}$, and $u = \frac{-1}{2N} \to u = \frac{1}{2N}$

By computing the integral with the new varibles

$$\int_{-\frac{1}{2N}}^{\frac{1}{2N}} u^k \, du = 2 \int_{0}^{\frac{1}{2N}} u^k \, du$$

k is even then u^k is none negative

and the integral can be simplified as

$$2 \left[\frac{u^{k+1}}{k+1} \right]_0^{\frac{1}{2N}} = \frac{2}{k+1} \left(\frac{1}{2N} \right)^{k+1} = \frac{1}{k+1} \cdot \frac{1}{2^k 2 N^{k+1}}$$

$$= \frac{|\Delta_i|^{k+1}}{2k(k+1)}$$

Finally $\int_{e_0}^{e_1} (t - t_i)^k \, dt = \frac{|\Delta_i|^{k+1}}{2k(k+1)}$

b. What are the optimal coefficients a_i and c_i that minimize the MSE of representing the entire signal using N intervals?

For finding the coefficients a_i and c_i that minimize the MSE we will use the MSE integral

$$E_i = \int_{e_0}^{e_1} \left(\phi(t) - \hat{\phi}(t)\right)^2 \quad \text{Where: } \hat{\phi}(t) = a(t-t_i) + C_i$$

with $t_i = \frac{2i-1}{2N} \rightarrow$ Average

$$E_i = \int_{e_0}^{e_1} \left(\phi(t) - a_i(t-t_i) * c_i\right)^2 \, dt =$$

$$\frac{\partial E_i}{\partial c_i} = -2\int_{e_0}^{e_1} \left(\phi(t) - a_i(t-t_i) - c_i\right)^2 dt = 0$$

$$c_i = \bar{\phi}_i - a_i \, \bar{t}_{i0}$$

$$\frac{\partial E_i}{\partial a_i} = -2\int_{e_0}^{e_1} \left(\phi(t) - a_i(t-t_i) - C_i\right) \, dt = 0$$

$\bar{\phi}$ is the avarge of $\phi(t)$ over Δi

$$d_i = \frac{\int_{c_0}^{c_1} (\phi(t) - \cancel{\bar{\phi}}_i)(t - t_i)\, dt}{\int_{c_0}^{c_1} -(t - t_i) - c_i \, dt}$$

c. Formulate the minimal MSE of representing the entire signal using N intervals.

$$\sum_{i=1}^{N} \left(\phi(t) - \phi'(t)\right)^2$$

$$\varphi(t)$$

$$MSE = \sum_{i=1}^{N} \int_{c_0}^{c_1} (\phi(t) - \bar{\phi}_i)^2 dt \quad \frac{\left(\int_{c_0}^{c_1} (\phi(t) - \bar{\phi}_i)(t - t_i)\, dt\right)^2}{\int_{c_0}^{c_1} (t - t_i)^2 dt}$$

$$\varphi'(t) \leftarrow$$

d. Compare the minimal MSE for using piecewise-linear approximation and the minimal MSE for using piecewise-constant approximation (as given in class – no need to develop it). Which MSE is lower? Mathematically justify your answer.

In piecewise constant aprroch each segment of the interval $[0,1]$ devided into N intervals that aproximated by the constant values c_i

$$MSE_{constant} = \sum_{i=1}^{N} \int_{c_0}^{c_1} (\phi(t) - c_i)^2 \, dt \;\Big|\; \text{where: } c_i = \bar{\phi}_i$$

the mean over

$$\Delta_i = \left[\frac{i-1}{N}, \frac{i}{N}\right]$$

piecewise approximation each interval

approximated by $a_i(t - t_i) + c_i$

$$MSE_{linear} = \sum_{i=1}^{N} \left(\int_{e_0}^{e_1} (\phi(t) - \bar{\phi})^2 \, dt - \left(\frac{\varphi(t)}{\varphi'_t} \right) \right)$$

the linear includes variance reduction term

$$RT = \left(\frac{\varphi(t)}{\varphi'(t)} \right)$$

→ squared, ≥ 0

→ integral of squares term, ≥ 0

ipynb?

Comparison of MSE for Piecewise-Linear and Piecewise-Constant Approximations

1. **Piecewise-Constant Approximation** where each interval uses a constant mean value.
2. **Piecewise-Linear Approximation** which uses a linear function within each interval.

Definitions

- **Piecewise-Constant MSE**:

$$\mathrm{MSE}_{\mathrm{constant}} = \sum_{i=1}^{N} \int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - \bar{\phi}_i)^2 \, dt$$

  Here, $\bar{\phi}_i$ is the mean of $\phi(t)$ over each interval Δ_i.

- **Piecewise-Linear MSE**:

$$\mathrm{MSE}_{\mathrm{linear}} = \sum_{i=1}^{N} \left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - \bar{\phi}_i)^2 \, dt - \frac{\left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - \bar{\phi}_i)(t - t_i) \, dt \right)^2}{\int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^2 \, dt} \right)$$

  where t_i is the center of each interval.

Mathematical Justification

To demonstrate that the MSE for the piecewise-linear approximation is always less than or equal to that for the piecewise-constant approximation:

1. **Definitions for Calculation**:
   - Let $V_i = \int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - \bar{\phi}_i)^2 \, dt$
   - Let $C_i = \left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - \bar{\phi}_i)(t - t_i) \, dt \right)^2$
   - Let $D_i = \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^2 \, dt$
2. **Using the Cauchy-Schwarz Inequality**:

$$\left(\int f g \, dx \right)^2 \leq \int f^2 \, dx \cdot \int g^2 \, dx$$

   By setting $f = \phi(t) - \bar{\phi}_i$ and $g = t - t_i$, we get:

$$C_i \leq V_i \cdot D_i$$

3. **Deriving the Inequality:**

$$\frac{C_i}{D_i} \leq V_i$$

   Therefore,

$$V_i - \frac{C_i}{D_i} \geq 0$$

   Leading to:

$$\mathrm{MSE}_{\mathrm{linear}} \leq \mathrm{MSE}_{\mathrm{constant}}$$

Conclusion

This derivation confirms that incorporating a linear component within each interval not only fits the average trend of the data but also adjusts to the variance within the interval, thus providing a better or at least equal approximation compared to the constant model.