HW 4 - Hints

HW4 HINTS AND FEEDBACK

EXERCISE 1.4

(1) We are interested in approximating

$$g = \underset{v \in \mathbb{P}_{\Lambda(w)}}{\arg\min} \| f - v \|^2_{L^2([0,1]^N)} \quad - (*)$$

for different test functions, f, that differ by regularity.

Here, we want to compute a polynomial approximation for f based on weighted discrete L^2 projection for $(*)$

$$\hat{f}_{\Lambda, M} = \underset{v \in \mathbb{P}_\Lambda([0,1]^N)}{\arg\min} \frac{1}{M} \sum_{k=1}^{M} \frac{\rho(y^{(k)})}{\hat{\rho}(y^{(k)})} |f - v|^2 (y^{(k)})$$

$- (**)$

Here,

- $\mathbb{P}_{\Lambda(\omega)} = \text{span}\left[\psi_{\underset{\sim}{p}}\, ,\ \underset{\sim}{p} \in \Lambda(\omega)\right]$

  $\underset{\sim}{p} \downarrow$ N - dimensional

- $\Lambda(\omega)$: index set

  That is $\Lambda = \{ \underset{\sim}{p}^{(1)}, \underset{\sim}{p}^{(2)}, \ldots \}$

  where $\underset{\sim}{p}^{(i)} = \{ p_1^{(i)}, \ldots, p_N^{(i)} \} \in \mathbb{N}^N$

  is a multi-index

- $$\psi_{\underset{\sim}{p}}(y) = \prod_{n=1}^{N} \psi_{p_n}(y_n)$$

  is the N- dimensional multi-variate polynomial basis

- $\psi_{p_n}(\cdot) \rightarrow$ polynomial of degree p_n in the n^{th} variable (out of N)

- M : number of samples used for estimation
- ρ : Original distribution (related to the choice of polynomial space for expansion)

For example, the Legendre polynomial forms an orthonormal basis in $L^2([-1,1])$ (i.e $\mathcal{U}[-1,1]$ distribution).

$\longrightarrow$ use this to find an orthonormal basis of $L^2([0,1])$

- Importance Sampling measure
  (Remember to motivate the use of IS)

Idea of L^2 regression is to represent $f(y)$ as linear combination of orthonormal polynomial basis functions in domain Γ.

$$\boxed{f(y) = \sum_{p \in \Lambda(w)} c_p \psi_p}$$

Choice of Index Set $\Lambda(w)$

Work with following index sets
(isotropic polynomial spaces)

a) Total degree (TD)

$$\Lambda^{TD}(w) = \left\{ \underline{p} \in \mathbb{N}^N ; \sum_{n=1}^{N} p_n \leq w \right\}$$

b) Hyperbolic Cross (HC)

$$\Lambda^{HC}(w) = \left\{ \underline{p} \in \mathbb{N}^N ; \prod_{n=1}^{N} (p_n + 1) \leq w + 1 \right\}$$

$\boxed{\text{Choice of } M \text{ based on stability}}$

Given $\Lambda(\omega)$, M is chosen to preserve stability of random discrete L^2 projection.

(see below)

NORMAL EQUATIONS

Solving $(**)$ is equivalent to expanding $\hat{f}_{\Lambda,M}$ onto orthonormal basis $\{\psi_p\}_{p\in\Lambda}$

$$\hat{f}_{\Lambda,M} = \sum_{p\in\Lambda} c_p \psi_p(y)$$

where the vector

$$\underline{c} = \{c_p\}_{p\in\Lambda} \quad \text{satisfies}$$

the normal equations

$$(D^T D)\,\underline{c} = D^T \underline{\phi}$$

where $\left(\underline{c} = \underset{\underline{v} \in \mathbb{R}^{|\Lambda|}}{\operatorname{argmin}} \| D\underline{v} - \underline{\phi} \|_2 \right)$

- $D \in \mathbb{R}^{M \times |\Lambda|}$ is design matrix

$$D_{i\underline{p}} = \psi_{\underline{p}}(y^{(i)})$$

with $\underline{p} \in \Lambda$ and $1 \leq i \leq M$

- $\underline{\phi} \Rightarrow \phi_i = f(y^{(i)})$, $1 \leq i \leq M$

Solution to normal equations is

$$\underline{c} = (D^T D)^{-1} D^T \underline{\phi}$$

If we define $G = \frac{1}{M} D^T D$, the stability of discrete least squares is related to $\| G^{-1} \|$

$\Rightarrow$ Check condition number of $(D^T D)$ for stability

Choose M such that $\text{cond}(D^T D)$ is bounded

- Try $\begin{cases} 1) \; M = c \cdot \#\Lambda \\ 2) \; M = c \cdot (\#\Lambda)^2 \end{cases}$ (* * *)

  with different values of c (1, 2, 3 etc...)

- Plot $\text{cond}(D^T D)$ with respect to level w of index set Λ.

- Plot $\left\| \frac{1}{M} D^T D - I \right\|$ for $M \to \infty$

  (you need a slightly different test when using importance sampling)

Choice of ρ and $\hat{\rho}$

CASE 1:

Use ρ uniform in $[0,1]$

$\Rightarrow$ expand function in Legendre polynomials

CASE 2:

For $\hat{\rho}$ $\begin{cases} \text{a) use uniform in } [0,1] \\ \text{b) use Chebyshev distribution} \end{cases}$

(also known as the "arcsine distribution")

⚠ Chebyshev distribution is in $[-1,1]$.
Map the points to suitable domain

- Compare CASE 1 and CASE 2 in terms of
  - i) Stability
  - ii) Error Convergence

Error Estimation

Error: $\| f - \hat{f}_{\Lambda,M} \|_{L^2([0,1]^N)}$

Use Cross-Validation

1. Use M_1 iid samples to approximate $\hat{f}_{\Lambda,M_1}$ (choose M_1 as in (***))

2. Independently use M_2 samples to estimate regression error

$$\boxed{\mathcal{E}_{cv}(M_2) = \frac{1}{M_2} \sum_{i=1}^{M_2} \frac{\rho(y^{(i)})}{\hat{\rho}(y^{(i)})} \left| f - \hat{f}_{\Lambda,M_1} \right|^2 (y^{(i)})}$$

- Plot $\mathcal{E}_{cv}$ as a function of M_2, for different choices of M_1 (linear/quadratic)

- Plot $\mathcal{E}_{cv}$ versus level of the index set w (for large enough M_2 (>30))

- Comment on the effect of dimension and regularity on the CV error

EXERCISE 1.4

② + ③

Trigonometric Approximation is based on

Exercise 1.4

2 + 3

Trigonometric Approximation is based on the orthonormal trigonometric basis functions in $L^2_\rho([0,2\pi])$

$$\phi_k(x) = e^{ikx}, \quad k \in \mathbb{Z}$$

⚠ k can also be negative

with $\rho \sim \mathcal{U}[0,2\pi]$

Recall!

$$K(\Lambda) = \sup_{y \in \Gamma} \left(\sum_{p \in \Lambda} |\psi_p(y)|^2 \right)$$

$$= \sup_{v \in \mathbb{P}_\Lambda} \frac{\|v\|_{L^\infty(\Gamma)}}{\|v\|_{L^2_\rho(\Gamma)}}$$

error split, see slide 663

$K(\Lambda)$ controls number of samples M to have optimal convergence of least square approximation.

① Unweighted Discrete Least Square

ρ : uniform and expansion in Legendre polynomials

$K(\Lambda) = (\#\Lambda)^2$ and $\boxed{M \propto (\#\Lambda)^2}$

② Weighted Discrete Least Square

$\begin{cases} \rho : \text{uniform ; expansion in Legendre polynomial} \\ \hat{\rho} : \text{Chebyshev distribution} \end{cases}$

dimension of problem

$K(\Lambda) \leq \min\left\{ (\#\Lambda)^{\log 3/\log 2} , 2^{N} \#\Lambda \right\}$

and $\boxed{M \propto 2^N . \#\Lambda}$

③ Unweighted Discrete Least Square for Periodic Functions

ρ : uniform and expansion in trigonometric basis

$K(\Lambda) = \#\Lambda$ and $\boxed{M \propto \#\Lambda}$

Adaptive Procedure to find Λ

DÖRFLER MARKING (see course slides for algorithm)

Reduced margin:

$$\mathcal{R}(\Lambda) = \left\{ \underline{p} : \underline{p} \notin \Lambda \text{ and } \forall j = 1, \dots, N : \; p_j \neq 0 \Rightarrow \left(\underline{p} - \underline{e}_j \in \Lambda \text{ or } \underline{p} + \underline{e}_j \in \Lambda \right) \right\}$$

Given
1) Multi-index set Λ
2) subset $R \subseteq \mathcal{R}(\Lambda)$
3) function $e : R \to \mathbb{R}$
4) parameter $\theta \in (0,1]$,

Dörfler marking computes set $F \subseteq R \subseteq \mathcal{R}(\Lambda)$ such that

$$\sum_{\underline{p} \in F} e(\underline{p})^2 \geq \theta \sum_{\underline{p} \in R} e(\underline{p})^2$$

The ORTHOGONAL MATCHING PURSUIT algorithm is used with Dörfler marking to adaptively choose Λ. (see course slides for algorithm)

In this algorithm, function $e(p)$ is either

1. Estimated coefficients $\underline{c} = \{c_p\}_{p \in \Lambda}$ or
2. Projected residual on p^{th} basis function

That is,

Let $r_k = f - \hat{f}_{\Lambda_k, M}$ be the residual at k^{th} iteration.

Then,

$$e(p) = (r_{k-1}, \psi_p)_M$$

$$= \frac{1}{M} \sum_{i=1}^{M} \left(f(y^{(i)}) - \Pi^M_{\Lambda_{k-1}}[f](y^{(i)}) \right) \psi_p(y^{(i)})$$

where the function $\Pi^M_{\Lambda_{k-1}}[f]$ is the result of the L^2 regression of f based on the index set Λ_{k-1} from the previous iteration.

f = Q(u_h)

- Plot $\mathcal{E}_{cv}$ versus $(\# \Lambda_k)$ for different choices of θ_1 and θ_2.

For instance:
$$\begin{cases} (\theta_1 = 0.5 \ ; \ \theta_2 = 0.2) \\ (\theta_1 = 0.9 \ ; \ \theta_2 = 0.7) \\ (\theta_1 = 0.2 \ ; \ \theta_2 = 0.5) \end{cases}$$

EXERCISE 2.2

① Use same guidelines as in previous tasks, where you work with $Q(u_h)$ instead of f.

Choice of h:

Choose h so that relative discretization error is 0.5%

② Bilevel L^2-regression approximation

Let $\Pi_{\Lambda}^{M}[f_h(y)]$ represent L^2 approx of $Q(u_h(y))$.

Bilevel approx,

$$S_1 f = \underbrace{\Pi_{\Lambda_0}^{M_0}[f_{h_0}(y)]}_{\mathrm{I}} + \underbrace{\Pi_{\Lambda_1}^{M_1}[(f_{h_1} - f_{h_0})(y)]}_{\mathrm{II}}$$

- $h_1 = h_0/2$
- $\Lambda_1 \subset \Lambda_0 \implies M_1 < M_0$
  (nested set)
- Apply previous guidelines to I and II separately to deduce M_1 and M_0

- Divide L^2 error contribution between I and II.

- Use cross-validation to deduce M_0, M_1 to achieve corresponding L^2-error

- Compare work needed to achieve L^2 error of 0.5 % using
  - i) bilevel
  - ii) single level

  discrete L^2-projection.

— X —