Lecture 2

17 August 2025 22:22

x is the designing parameter

Energy :- (CV)

Def - $E(x) = D(x) + \lambda R(x)$

E → Energy, $D(x)$ → Data fidelity, $R(x)$ → Regularizer

λ is some scalar
$\lambda > 0$

$D(x)$ measures how well x explains the observations

For example, $D(x) = -\log(\underbrace{\mathbb{P}(\text{data}|x)}_{\text{Posterior probability}})$

(probability or distribution of data after observation (of x))

In terms of $\mathbb{P}(x|\text{data})$ we write $\mathbb{P}(\text{data}|x) = \dfrac{\mathbb{P}(x|\text{data})\,\mathbb{P}(\text{data})}{\mathbb{P}(x)}$

$\mathbb{P}(\text{data}|x)$ → posterior; $\mathbb{P}(x|\text{data})$ → Likelihood; $\mathbb{P}(x)$ → evidence; $\mathbb{P}(\text{data})$ → Prior

comes from prior

$R(x)$ gives the structure or distribution we get for x.

For example, $-\log(\mathbb{P}(x))$

The work of λ to make a stable relation, i.e., get a sweet spot for data fit and prior distribution

Our OP :- $\underset{x \in S}{\arg\min}(E(x)) = \underset{x \in S}{\arg\min}(D(x) + \lambda R(x))$

For example we get, for $\lambda = 1$, $\underset{x \in S}{\arg\min}\left(-\log(\mathbb{P}(\text{data}|x)) - \log \mathbb{P}(x)\right)$

$$\Rightarrow \underset{x \in S}{\arg\min}\left(-\log\left(\mathbb{P}(\text{data}|x)\cdot\mathbb{P}(x)\right)\right)$$

$$= \underset{x \in S}{\arg\min}\left(-\log\left(\mathbb{P}(x|\text{data})\,\mathbb{P}(\text{data})\right)\right)$$

$$= \underset{x \in S}{\arg\min}\left(-\log \mathbb{P}(x|\text{data})\right)$$

$$= \underset{x \in S}{\arg\max}\left(\log \mathbb{P}(x|\text{data})\right)$$

Some examples of D :-

1) Gaussian (noise) :- $D(x) = \frac{1}{2\sigma^2} \|\varepsilon\|_2^2 = \frac{1}{2\sigma^2} \| Mx - f\|_2^2 + \text{const}$

from $f = Mx + \varepsilon$, $\varepsilon \in \mathcal{N}(0, \sigma^2 I_m)$

$\hookrightarrow$ transformation matrix

all f_i's, x_i's are independent distributed and identically

Likelihood $= \mathbb{P}(f|x) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{1}{2\sigma^2} \| Mx - f\|_2^2\right)$

$D(x) = -\log\left(\mathbb{P}(f|x)\right) = \log\left(\frac{1}{(2\pi\sigma^2)^{m/2}}\right) - \frac{1}{2\sigma^2} \| Mx - f\|_2^2$

$= -\frac{1}{2\sigma^2} \|Mx - f\|_2^2 + \text{const.}$

2) Poisson noise :- $f(x) = Mx + \varepsilon$, ε is poission distribution

$D(x) = \sum_i \left((Mx)_i - f_i \log (Mx)_i \right) + \text{const}$

$\varepsilon \sim \frac{e^{-\sum \lambda_i} \left(\prod \lambda_i^{k_i}\right)}{\prod (k_i)!}$, $\quad \varepsilon_i \sim \frac{e^{-\lambda_i} \lambda_i^{k_i}}{k_i!}$ $\quad \varepsilon_i$'s are independent

Note :- we will come to outliers and robustness later

Some examples of R :-

1) Tikhonor :- $R(x) = \frac{1}{2} \|\nabla_x g\|_2^2$ (used generally for smoothening)

(note :- to check multivariate transformation of g)

(note :- to check multivariate transformation of g)

2) Total Variance (TV) :- $R(x) = \|\nabla_x g\|_{2,1} = \int_{\Omega} \|\nabla_x g\|_2 \, dz$

(piecewise functions)

$\ell_{2,1}$ norm (it is a matrix norm)

$\Omega \Rightarrow \Omega \in \mathbb{R}^d \to \mathbb{R}$

For example, let $X \in \mathbb{R}^{m \times n}$ is a matrix

$$\|X\|_{2,1} = \sum_{i=1}^{m} \|X_i\|_2$$

X_i is the ith row