Regression based on the $\mathcal{K}$-divergence:

Consider the following presumed regression model:

$$\boxed{y = \varphi(\mathbf{x};\boldsymbol{\tau}) + \xi}$$

Where $\boldsymbol{\tau}$ is the vector parameter of interest of the target function $\varphi(\cdot;\cdot)$.

The noise is assumed to obey a Gaussian distribution $\xi \sim N(0,\sigma^2)$. We define $\boldsymbol{\theta} \in \mathbb{R}^{M+1}$ as $\boldsymbol{\theta} \triangleq \left[\boldsymbol{\tau}^T, \sigma^2\right]^T$. By this assumption, the joint probability of the inputs and outputs $\mathbf{z} = \left[\mathbf{x}^T, y\right]^T$:

$$\boxed{f_{y|\mathbf{x}}(y \mid \mathbf{x};\boldsymbol{\theta}) = \phi(y;\varphi(\mathbf{x};\boldsymbol{\tau}),\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\varphi(\mathbf{x};\boldsymbol{\tau}))^2}{\sigma^2}}}$$

Where the function $\phi\left(x;\mu,\sigma^2\right) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is a Gaussian p.d.f.

We estimate the parameter $\boldsymbol{\theta}$ by minimizing the penalized loss:

$$\boxed{L_{\mathbf{h},\lambda}^{(R)}\left(\boldsymbol{\theta}\right) = C_h^{(R)}\left(\boldsymbol{\theta}\right) + \lambda P(\boldsymbol{\tau})}$$

Where $P(\boldsymbol{\tau})$ is the penalty function of $\boldsymbol{\tau}$ and λ is a regularization parameter.

The non-penalized $\mathcal{K}$-loss is given by:

$$\boxed{J_h^{(R)}\left(\boldsymbol{\theta}\right) \triangleq \sum_{n=1}^{N} w\left(\mathbf{x}_n, y_n;\mathbf{h}\right)\log f_{\boldsymbol{\theta}}(y_n \mid \mathbf{x}_n) - \log \hat{u}(\mathbf{h},\boldsymbol{\theta})}$$

We choose the kernel function as a spherical Gaussian function that follows:

$$\boxed{\begin{array}{l} K_{\mathbf{h}}\left(\mathbf{r}\right) \triangleq = \frac{1}{\left(2\pi h_x^2\right)^{\frac{p}{2}}} \exp\left(-\frac{\left\|\mathbf{s}\right\|^2}{2h_x^2}\right) \frac{1}{\left(2\pi h_y^2\right)^{\frac{1}{2}}} \exp\left(-\frac{t^2}{2h_y^2}\right) = \\ \underbrace{\phi\left(\mathbf{s};\mathbf{0},h_x^2\mathbf{I}_p\right)}_{\triangleq K_{h_{\mathbf{x}}}(\mathbf{s})} \underbrace{\phi\left(t;0,h_y^2\right)}_{\triangleq K_{h_y}(t)} = K_{h_{\mathbf{x}}}\left(\mathbf{s}\right) K_{h_y}\left(t\right) \\ \mathbf{h} \triangleq [h_x, h_y]^T,\ \mathbf{r} = [\mathbf{s}^T, t]^T \end{array}}$$

By this choice, we note that the weighting function:

$$\boxed{\begin{gathered}w(\mathbf{s},t;\mathbf{h})=\frac{\hat{g}_{\mathbf{x},y}(\mathbf{s},t;\mathbf{h})-\frac{1}{N}K_{h_{\mathbf{x}}}(\mathbf{0})K_{h_y}(0)}{\sum_{m=1}^{N}\left(\hat{g}_{\mathbf{x},y}(\mathbf{x}_m,y_m;\mathbf{h})-\frac{1}{N}K_{h_{\mathbf{x}}}(\mathbf{0})K_{h_y}(0)\right)}=\\ \frac{\sum_{n=1}^{N}\left(\exp\left(-\frac{\|\mathbf{s}-\mathbf{x}_n\|^2}{2h_x^2}\right)\exp\left(-\frac{(t-y_n)^2}{2h_y^2}\right)\right)-\frac{1}{N}}{\sum_{m=1}^{N}\left(\sum_{n=1}^{N}\left(\exp\left(-\frac{\|\mathbf{x}_m-\mathbf{x}_n\|^2}{2h_x^2}\right)\exp\left(-\frac{(y_m-y_n)^2}{2h_y^2}\right)\right)-\frac{1}{N}\right)}\end{gathered}}$$

and:

$$\hat{u}(\mathbf{h},\boldsymbol{\theta})=\frac{1}{N-1}\sum_{n=1}^{N}\int_{\mathbb{R}}\bar{g}(\mathbf{x}_n,t;\mathbf{h})f_{\boldsymbol{\theta}}(t\mid\mathbf{x}_n)d\lambda(t)$$

where:

$$\begin{aligned}&\int_{\mathbb{R}}\bar{g}(\mathbf{x}_n,t;\mathbf{h})f_{\boldsymbol{\theta}}(t\mid\mathbf{x}_n)d\lambda(t)=\int_{\mathbb{R}}\left(\frac{1}{N}\sum_{m=1}^{N}K_{h_{\mathbf{x}}}(\mathbf{x}_n-\mathbf{x}_m)K_{h_y}(t-y_m)\right)\phi\left(t;\varphi(\mathbf{x}_n;\boldsymbol{\tau}),\sigma^2\right)d\lambda(t)=\\ &\frac{1}{N}\sum_{m=1}^{N}K_{h_x}(\mathbf{x}_n-\mathbf{x}_m)\int_{\mathbb{R}}\phi\left(t;y_m,h_y^2\right)\phi\left(t;\varphi(\mathbf{x}_k;\boldsymbol{\tau}),\sigma^2\right)d\lambda(t)-\frac{1}{N}K_h^{(\mathbf{x})}(\mathbf{0})\int_{\mathbb{R}}\phi\left(t;y_n,h_y^2\right)\phi\left(t;\varphi(\mathbf{x}_n;\boldsymbol{\tau}),\sigma^2\right)d\lambda(t)=\\ &\frac{1}{N}\sum_{m=1}^{N}K_{h_x}(\mathbf{x}_n-\mathbf{x}_m)\phi\left(y_m;\varphi(\mathbf{x}_n;\boldsymbol{\tau}),h_y^2+\sigma^2\right)-\frac{1}{N}K_{h_x}(\mathbf{0})\phi\left(y_n;\varphi(\mathbf{x}_n;\boldsymbol{\tau}),h_y^2+\sigma^2\right)\\ &=\frac{1}{N}\sum_{m\neq n}K_{h_x}(\mathbf{x}_n-\mathbf{x}_m)\phi\left(y_m;\varphi(\mathbf{x}_n;\boldsymbol{\tau}),h_y^2+\sigma^2\right)\end{aligned}$$

and thus:

$$\boxed{\hat{u}(\mathbf{h},\boldsymbol{\theta})=\frac{1}{N(N-1)}\sum_{n=1}^{N}\sum_{m\neq n}K_{h_x}(\mathbf{x}_n-\mathbf{x}_m)\phi\left(y_m;\varphi(\mathbf{x}_k;\boldsymbol{\tau}),h_y^2+\sigma^2\right)}$$

Derivatives calculations for Gradient ascent:

We will find the optimal $\boldsymbol{\theta}$ by Gradient ascent approach:

$$\boxed{\begin{aligned}&\boldsymbol{\theta}_{i+1}=\boldsymbol{\theta}_i+\beta\nabla_{\boldsymbol{\theta}}L_{\mathbf{h},\lambda}^{(R)}\left(\boldsymbol{\theta}_i\right)=\boldsymbol{\theta}_i+\beta\left(\nabla_{\boldsymbol{\theta}}C_{\mathbf{h},\lambda}^{(R)}\left(\boldsymbol{\theta}_i\right)+\lambda\nabla_{\boldsymbol{\tau}}P\left(\boldsymbol{\tau}_i\right)\right)\\&\boldsymbol{\theta}_i\triangleq\left[\boldsymbol{\tau}_i^T,\sigma_i^2\right]\end{aligned}}$$

where β is the step size. We now calculate the derivative $\nabla_{\boldsymbol{\theta}}C_h^{(R)}\left(\boldsymbol{\theta}\right)$.

$$\boxed{\begin{aligned}&\nabla_{\boldsymbol{\theta}}C_h^{(R)}\left(\boldsymbol{\theta}\right)=\sum_{n=1}^{N}w(\mathbf{x}_n,y_n;\mathbf{h})\mathbf{q}\left(\mathbf{x}_n,y_n;\boldsymbol{\theta}\right)-\frac{\boldsymbol{\psi}\left(\mathbf{h},\boldsymbol{\theta}\right)}{U\left(\mathbf{h},\boldsymbol{\theta}\right)}\\[2ex]&\boldsymbol{\psi}\left(\mathbf{h},\boldsymbol{\theta}\right)\triangleq\sum\nolimits_{n=1}^{N}\sum_{m\neq n}K_{h_x}(\mathbf{x}_k-\mathbf{x}_m)\nabla_{\boldsymbol{\theta}}\phi\left(y_m;\varphi\left(\mathbf{x}_n;\boldsymbol{\tau}\right),h_y^2+\sigma^2\right)=\\&\sum\nolimits_{n=1}^{N}\sum_{m\neq k}K_{h_x}^{(\mathbf{x})}(\mathbf{x}_k-\mathbf{x}_m)\phi\left(y_m;\varphi\left(\mathbf{x}_n;\boldsymbol{\eta}\right),h_y^2+\sigma^2\right)\mathbf{q}\left(\mathbf{x}_n,y_m;\tilde{\boldsymbol{\theta}}_h\right)\\[2ex]&\tilde{\boldsymbol{\theta}}_h\triangleq\left[\boldsymbol{\eta}^T,h_y^2+\sigma^2\right]^T\\[2ex]&U\left(\mathbf{h},\boldsymbol{\theta}\right)=\sum\nolimits_{n=1}^{N}\sum_{m\neq n}K_{h_x}^{(\mathbf{x})}(\mathbf{x}_k-\mathbf{x}_m)\phi\left(y_m;\varphi\left(\mathbf{x}_n;\boldsymbol{\tau}\right),h_y^2+\sigma^2\right)\end{aligned}}$$

I now calculate the derivatives:

$$\boxed{\mathbf{q}\left(\mathbf{s},t;\boldsymbol{\theta}\right)\triangleq\nabla_{\boldsymbol{\theta}}\log\phi\left(t;\varphi\left(\mathbf{s};\boldsymbol{\tau}\right),\sigma^2\right)=\begin{bmatrix}\nabla_{\boldsymbol{\eta}}\log\phi\left(t;\varphi\left(\mathbf{s};\boldsymbol{\tau}\right),\sigma^2\right)\\\dfrac{\partial\log\phi\left(t;\varphi\left(\mathbf{s};\boldsymbol{\tau}\right),\sigma^2\right)}{\partial\sigma^2}\end{bmatrix}}$$

First, by the chain-rule we note that:

$$\boxed{\nabla_{\boldsymbol{\tau}}\log\phi\left(t;\varphi\left(\mathbf{s};\boldsymbol{\tau}\right),\sigma^2\right)=\left.\frac{\partial\log\phi\left(t;\mu,\sigma^2\right)}{\partial\mu}\right|_{\mu=\varphi\left(\mathbf{s};\boldsymbol{\tau}\right)}\nabla_{\boldsymbol{\tau}}\varphi\left(\mathbf{s};\boldsymbol{\tau}\right)}$$

where:

$$\boxed{\begin{aligned}&\log\phi\left(t;\mu,\sigma^2\right)=-\frac{1}{2}\log\sigma^2-\frac{(t-\mu)^2}{2\sigma^2}-\frac{1}{2}\log2\pi\\&\frac{\partial\log\phi\left(t;\mu,\sigma^2\right)}{\partial\mu}=\frac{(t-\mu)}{\sigma^2}\end{aligned}}$$

Lastly, I calculate the derivative w.r.t the noise variance, where I use the auxiliary formula:

$$\boxed{\frac{\partial \log \phi\left(t;\mu,\sigma^2\right)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} = \frac{1}{2\sigma^2}\left(\frac{(x-\mu)^2}{\sigma^2} - 1\right)}$$

To summarize:

$$\boxed{\nabla_{\boldsymbol{\theta}} C_{\mathbf{h}}^{(R)}\left(\boldsymbol{\theta}\right) \triangleq \begin{bmatrix} \nabla_{\boldsymbol{\tau}} C_{\mathbf{h}}^{(R)}\left(\boldsymbol{\theta}\right) \\ \dfrac{\partial C_{\mathbf{h}}^{(R)}\left(\boldsymbol{\theta}\right)}{\partial \sigma^2} \end{bmatrix}}$$

$$\boxed{\nabla_{\boldsymbol{\tau}} J_{h}^{(R)}\left(\boldsymbol{\theta}\right) = \sum_{n=1}^{N} \tilde{\lambda}_{\mathbf{h}}\left(\mathbf{x}_n, y_n;\boldsymbol{\theta}\right) \nabla_{\boldsymbol{\eta}} \varphi\left(\mathbf{x}_n;\boldsymbol{\eta}\right)}$$

$$\tilde{\lambda}_{\mathbf{h}}\left(\mathbf{x}_n, y_n;\boldsymbol{\theta}\right) \triangleq \frac{1}{\sigma^2} w(\mathbf{x}_n, y_n;\mathbf{h})\left(y_n - \varphi(\mathbf{x}_n;\boldsymbol{\tau})\right) - \frac{1}{h_y^2+\sigma^2}\zeta_{\mathbf{h}}(\mathbf{x}_n, y_n;\boldsymbol{\theta})$$

$$\zeta_{\mathbf{h}}(\mathbf{x}_n, y_n;\boldsymbol{\theta}) \triangleq \frac{\sum_{m\neq n} K_{h_x}^{(\mathbf{x})}(\mathbf{x}_n - \mathbf{x}_m)\phi\left(y_m;\varphi\left(\mathbf{x}_n;\boldsymbol{\tau}\right), h_y^2+\sigma^2\right)\left(y_m - \varphi(\mathbf{x}_n;\boldsymbol{\tau})\right)}{\sum_{k=1}^{N}\sum_{j\neq k} K_{h_x}^{(\mathbf{x})}(\mathbf{x}_k - \mathbf{x}_j)\phi\left(y_j;\varphi\left(\mathbf{x}_k;\boldsymbol{\tau}\right), h_y^2+\sigma^2\right)} =$$

$$\sum_{m=1}^{N}\left(\gamma_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n;\boldsymbol{\theta})\left(y_m - \varphi(\mathbf{x}_n;\boldsymbol{\eta})\right) - \frac{1}{N}\gamma_{\mathbf{h}}(\mathbf{z}_n, \mathbf{z}_n;\boldsymbol{\theta})\left(y_n - \varphi(\mathbf{x}_n;\boldsymbol{\eta})\right)\right)$$

$$\gamma_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n;\boldsymbol{\theta}) \triangleq \frac{t_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n;\boldsymbol{\theta})}{\sum_{k=1}^{N} s_{\mathbf{h}}(\mathbf{z}_k;\boldsymbol{\theta})}$$

$$s_{\mathbf{h}}(\mathbf{z}_n;\boldsymbol{\theta}) \triangleq \sum_{m=1}^{N}\left(t_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n;\boldsymbol{\theta}) - \frac{1}{N} t_{\mathbf{h}}(\mathbf{z}_n, \mathbf{z}_n;\boldsymbol{\theta})\right)$$

$$t_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n;\boldsymbol{\theta}) \triangleq e^{-\frac{\left\|\mathbf{x}_m - \mathbf{x}_n\right\|^2}{h_x^2}} e^{-\frac{\left(y_m - \varphi(\mathbf{x}_n;\boldsymbol{\eta})\right)^2}{h_y^2+\sigma^2}}$$

$$\mathbf{z}_n \triangleq \left[\mathbf{x}_n^T, y_n\right]^T, \ \forall n = 1,\ldots,N$$

and

$$\boxed{\frac{\partial C_{h}^{(R)}\left(\boldsymbol{\theta}\right)}{\partial \sigma^2} = \sum_{n=1}^{N} \tilde{\xi}_h\left(\mathbf{x}_n, y_n;\boldsymbol{\theta}\right)}$$

$$\tilde{\xi}_{\mathbf{h}}\left(\mathbf{x}_n, y_n; \boldsymbol{\theta}\right) \triangleq \frac{1}{2\sigma^4} w(\mathbf{x}_n, y_n; \mathbf{h})\left(y_n - \varphi(\mathbf{x}_n; \boldsymbol{\eta})\right)^2 - \frac{1}{2\sigma^2} - \frac{1}{2\left(h_y^2 + \sigma^2\right)^2}\psi_{\mathbf{h}}(\mathbf{x}_n, y_n; \boldsymbol{\theta}) + \frac{1}{2\left(h_y^2 + \sigma^2\right)} =$$

$$\frac{1}{2\sigma^4} w(\mathbf{x}_n, y_n; \mathbf{h})\left(y_n - \varphi(\mathbf{x}_n; \boldsymbol{\eta})\right)^2 - \frac{1}{2\left(h_y^2 + \sigma^2\right)^2}\psi_{\mathbf{h}}(\mathbf{x}_n, y_n; \boldsymbol{\theta}) - \frac{h_y^2}{2\sigma^2\left(h_y^2 + \sigma^2\right)}$$

$$\psi_{\mathbf{h}}(\mathbf{x}_n, y_n; \boldsymbol{\theta}) \triangleq \frac{\sum_{m \neq n} K_{h_x}^{(\mathbf{x})}(\mathbf{x}_n - \mathbf{x}_m)\phi\left(y_m; \varphi(\mathbf{x}_n; \boldsymbol{\tau}), h_y^2 + \sigma^2\right)\left(y_m - \varphi(\mathbf{x}_n; \boldsymbol{\tau})\right)^2}{\sum_{k=1}^{N}\sum_{j \neq k} K_{h_x}^{(\mathbf{x})}(\mathbf{x}_k - \mathbf{x}_j)\phi\left(y_j; \varphi(\mathbf{x}_k; \boldsymbol{\tau}), h_y^2 + \sigma^2\right)} =$$

$$\sum_{m=1}^{N}\left(\gamma_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n; \boldsymbol{\theta})\left(y_m - \varphi(\mathbf{x}_n; \boldsymbol{\eta})\right)^2 - \frac{1}{N}\gamma_{\mathbf{h}}(\mathbf{z}_n, \mathbf{z}_n; \boldsymbol{\theta})\left(y_n - \varphi(\mathbf{x}_n; \boldsymbol{\eta})\right)^2\right)$$

$$\gamma_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n; \boldsymbol{\theta}) \triangleq \frac{t_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n; \boldsymbol{\theta})}{\sum_{k=1}^{N} s_{\mathbf{h}}(\mathbf{z}_k; \boldsymbol{\theta})}$$

$$s_{\mathbf{h}}(\mathbf{z}_n; \boldsymbol{\theta}) \triangleq \sum_{m=1}^{N}\left(t_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n; \boldsymbol{\theta}) - \frac{1}{N}t_{\mathbf{h}}(\mathbf{z}_n, \mathbf{z}_n; \boldsymbol{\theta})\right)$$

$$t_{\mathbf{h}}(\mathbf{z}_m, \mathbf{z}_n; \boldsymbol{\theta}) \triangleq e^{-\frac{\left\|\mathbf{x}_m - \mathbf{x}_n\right\|^2}{h_x^2}} e^{-\frac{\left(y_m - \varphi(\mathbf{x}_n; \boldsymbol{\tau})\right)^2}{h_y^2 + \sigma^2}}$$

$$\mathbf{z}_n \triangleq \left[\mathbf{x}_n^T, y_n\right]^T, \ \forall n = 1, \ldots, N$$