Exercise 2.1. Gradient descent of FoE.

Show that the gradient descent of E_{θ}^{FoE} is as ch:

$$u^{t+1} = u^t - \eta\left(\frac{1}{\sigma^2}(u^t - v) + \sum_{i=1}^{N} \bar{k}_i * \varphi_i'(k_i * u^t)\right), \quad t = 0, \dots, T.$$

Recall that a gradient descent is an iterative minimization method with the following update equation:

$$u^{t+1} = u^t - \eta \nabla E_{\theta}^{FoE}(u_t, v).$$

We have to compute ∇E_{θ}^{FoE}.

$$E_{\theta}^{FoE} = \frac{1}{2\sigma^2}\|u - v\|^2 + \sum_{x\in\Omega}\sum_{i=1}^{N} \varphi_i(k_i * u(x)) + C.$$

Let's start part by part.

$$\nabla\left(\frac{1}{2\sigma^2}\|u - v\|^2\right) = \frac{1}{\sigma^2}(u - v)$$

$$\nabla\left(\sum_{x\in\Omega}\sum_{i=1}^{N} \varphi_i(k_i * u(x))\right) = \sum_{x\in\Omega}\sum_{i=1}^{N} \nabla\varphi_i(k_i * u(x)) = \sum_{x\in\Omega}\sum_{i=1}^{N} \nabla(k_i * u(x)) \cdot \varphi_i'(k_i * u(x)) = \sum_{x\in\Omega}\sum_{i=1}^{N} \varphi_i'(k_i * u(x)) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ k_i[x] \\ \vdots \\ k_i[1] \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Next, we will change summation order.

$$\sum_{i=1}^{N}\sum_{x\in\Omega} \begin{pmatrix} 0 \\ \vdots \\ k_i[x] \\ \vdots \\ k_i[1] \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varphi_i'(k_i * u(x)) = \sum_{i=1}^{N} \bar{k}_i * \varphi_i'(k_i * u(x)).$$

$$\nabla C = 0 \implies \nabla E_{\theta}^{FoE} = \frac{1}{\sigma^2}(u - v) + \sum_{i=1}^{N} \bar{k}_i * \varphi_i'(k_i * u(x)) \implies$$

$$\implies u^{t+1} = u^t - \eta\left(\frac{1}{\sigma^2}(u - v) + \sum_{i=1}^{N} \bar{k}_i * \varphi_i'(k_i * u(x))\right)$$

Exercise 2.2. Maximum likelihood as density matching.

The Kullback-Leibler divergence (also called relative entropy) is a measure of how one probability distribution q differs from a second reference probability distribution p, given by:

$$KL(p(x)\|q(x)) = \int \log\left(\frac{p(x)}{q(x)}\right) p(x)\,dx$$

We assume that data distributed according to $p(x)$. Consider the limit of the log-likelihood when $m \to +\infty$:

$$\mathcal{L}_\infty(\theta) = E\{\log p_\theta(x)\} = \int \log p_\theta(x)\, p(x)\,dx = $$
$$= \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} \log p_\theta(x_j).$$

Show that maximizing the ~~expected~~ log-likelihood $\mathcal{L}_\infty$ minimizes

$$KL(p(x)\|p_\theta(x)) = \int \log\left(\frac{p(x)}{p_\theta(x)}\right) p(x)\,dx$$

$$\max_\theta \mathcal{L}_\infty(\theta) = \max_\theta \int \log(p_\theta(x))\, p(x)\,dx = \min_\theta -\int \log(p_\theta(x))\, p(x)\,dx$$

Also, $\min_\theta KL(p(x)\|p_\theta(x)) = \min_\theta \int \log\left(\frac{p(x)}{p_\theta(x)}\right) p(x)\,dx =$
$= \min_\theta \int \log(p(x))\,p(x) - \log(p_\theta(x))\,p(x)\,dx$. The first part is independent of θ, so it is not involved in minimization, we have $\min_\theta -\int \log(p_\theta(x))\,p(x)\,dx + \int \log(p(x))\,p(x)\,dx \Rightarrow$ We have to solve the $\min_\theta -\int \log(p_\theta(x))\, p(x)\,dx$ which is the same as we got before, so maximizing $\mathcal{L}_\infty(\theta)$ minimizes $KL(p(x)\|p_\theta(x))$