$\underset{\sim}{\theta}_N = (\theta_1, \ldots, \theta_N)$ model parameter. $\underset{\sim}{x} = (x_1, \ldots, x_p)$: data.

posterior: $$\Pi_n(\underset{\sim}{\theta}_N) = \frac{e^{l_n(\underset{\sim}{\theta}_N)} \cdot \pi(\underset{\sim}{\theta}_N)}{\int e^{l_n(\underset{\sim}{\theta}_N)} \cdot \pi(\underset{\sim}{\theta}_N) \, d\underset{\sim}{\theta}_N}.$$

$$l_n(\underset{\sim}{\theta}_N) = \sum_{i=1}^{p} l_n(x_i \mid \underset{\sim}{\theta}_N),$$

and $$\pi(\underset{\sim}{\theta}_N) = \prod_{i=1}^{N} \pi_0(\theta_i).$$

$$\Rightarrow \quad \Pi_n(\underset{\sim}{\theta}_N) = \frac{\exp\left(\sum_{i=1}^{p} l_n(x_i \mid \underset{\sim}{\theta}_N)\right) \cdot \prod_{i=1}^{N} \pi_0(\theta_i)}{\int e^{l_n(\underset{\sim}{\theta}_N)} \cdot \pi(\underset{\sim}{\theta}_N) \, d\underset{\sim}{\theta}_N}.$$

- De Bortoli et al (2020): "Quantitative Propagation of Chaos for SGD in Wide NN".

  - Mean-Field Approximation, solution of a mean-field McKean-Vlasov equation.

  - stepsize in SGD depending on the number of hidden units leads to particle system with two possible mean-field behaviours
    1. Mean-field ODE
    2. McKean-Vlasov diffussion.

step-size.

- $W_{n+1}^{1:N} = W_n^{1:N} - \boxed{\gamma N^{\beta} (n + \gamma_{\alpha,\beta}(N)^{-1})^{-\alpha}} \nabla \hat{R}^N (W_n^{1:N}, X_n, Y_n)$

  $\beta \in [0,1]$, $\alpha \in [0,1)$, $\gamma_{\alpha,\beta}(N) = \gamma^{1/(1-\alpha)} N^{(\beta-1)/(1-\alpha)}$.

- Mei et al. (2018): "A Mean Field View of the Landscape of Two-Layer NN".

  - optimize a non-convex high-dimensional object (risk function).

  - SGD dynamics is captured by non-linear PDE: distributional dynamics (DD).

  - PDE converges to a near global optimum, when initialized with a bounded density.

- Chizat and Bach (2018):

  min convex function, 2 layer NN, over-parameterized, non-convex gradient flow.

  particle gradient flow. when initialized correctly and in the many-particle limit, this gradient flow, although non-convex, converges to global minimizers.

- Rotskoff et al. (2019):

  - relies on taking a "mean-field" limit in which the number of parameters n tends to ∞.

    (Mean-field)

  - Gradient descent obeys a deterministic PDE that converges to a globally optimal solution.

- Tzen and Raginsky (2019):

  Diffusion Limit, discretize-then-differentiate approach,

  Variational inference in neural SDEs.