



hw13.pdf
PDF 217.32KB

Do all problems in the attached pdf. Solve them one by one with detail explanations.

⊗ 已思考 (用时 382 秒) >

Problem 1: DDPM/DDIM Fun: From a Gaussian

(a) Distribution of X_1

The forward process adds independent Gaussian noise of variance Δt at each of $T = 1/\Delta t$ steps. Starting from $X_0 \sim \mathcal{N}(0, \sigma^2)$, the total added noise is $\mathcal{N}(0, 1)$. Thus,

$$X_1 = X_0 + \mathcal{N}(0, 1) \sim \mathcal{N}(0, \sigma^2 + 1).$$

(b) Marginal distributions of $X_{t-\Delta t}$ and X_t

At time t , the accumulated noise variance is t . Hence,

$$X_t = X_0 + \mathcal{N}(0, t) \sim \mathcal{N}(0, \sigma^2 + t),$$

$$X_{t-\Delta t} = X_0 + \mathcal{N}(0, t - \Delta t) \sim \mathcal{N}(0, \sigma^2 + t - \Delta t).$$

(c) Approximate conditional variance

Given $(X_{t-\Delta t} | X_t = x_t) \sim \mathcal{N}\left(\frac{\sigma^2 + t - \Delta t}{\sigma^2 + t}x_t, \frac{(\sigma^2 + t - \Delta t)\Delta t}{\sigma^2 + t}\right)$.

Since $\Delta t \ll \sigma^2 \ll 1$, we have $\sigma^2 + t - \Delta t \approx \sigma^2 + t$. Thus,

$$\text{Var} \approx \frac{(\sigma^2 + t)\Delta t}{\sigma^2 + t} = \Delta t.$$

(d) Reverse process without stochastic noise

Applying only the conditional mean backward:

$$\hat{X}_{t-\Delta t} = \frac{\sigma^2 + t - \Delta t}{\sigma^2 + t} \hat{X}_t.$$

Starting with $\hat{X}_1 \sim \mathcal{N}(0, 1)$, after T steps:

$$\hat{X}_0 = \hat{X}_1 \prod_{k=0}^{T-1} \frac{\sigma^2 + 1 - (k+1)\Delta t}{\sigma^2 + 1 - k\Delta t} = \hat{X}_1 \cdot \frac{\sigma^2}{\sigma^2 + 1}.$$

Thus,

$$\hat{X}_0 \sim \mathcal{N}\left(0, \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2\right) \approx \mathcal{N}(0, \sigma^4).$$

(e) Reverse process with added noise

Given:

$$\text{Var}(\hat{X}_0) = \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2 + \Delta t \sum_{k=0}^{T-1} \left(\frac{\sigma^2}{\sigma^2 + k\Delta t}\right)^2.$$

As $\Delta t \rightarrow 0$, the sum becomes an integral:

$$\text{Var}(\hat{X}_0) \approx \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2 + \int_0^1 \left(\frac{\sigma^2}{\sigma^2 + t}\right)^2 dt.$$

Compute the integral:

$$\int_0^1 \frac{\sigma^4}{(\sigma^2 + t)^2} dt = \sigma^4 \left[-\frac{1}{\sigma^2 + t} \right]_0^1 = \sigma^2 - \frac{\sigma^4}{\sigma^2 + 1}.$$

Thus,

$$\text{Var}(\hat{X}_0) \approx \left(\frac{\sigma^2}{\sigma^2 + 1}\right)^2 + \sigma^2 - \frac{\sigma^4}{\sigma^2 + 1} = \sigma^2 - \frac{\sigma^6}{(\sigma^2 + 1)^2} \approx \sigma^2.$$

(f) Training a neural network to estimate $g(x_t, t)$

- **Inputs:** Noisy sample x_t and time t .
- **Batch generation:** For each clean sample s_i from the dataset, sample $t \sim U[0, 1]$ and $z \sim \mathcal{N}(0, 1)$. Compute $x_t = s_i + \sqrt{t}z$. The target is $y = \frac{\sigma^2}{\sigma^2 + t}x_t$.
- **Loss:** Mean squared error between the network output and y .

(g) Approximate DDIM step

For $\Delta t \ll t$, approximate $\eta(t, \Delta t) = \frac{\sqrt{t}}{\sqrt{t-\Delta t} + \sqrt{t}} \approx \frac{1}{2}$. The deterministic DDPM step is $-\frac{\Delta t}{\sigma^2 + t}x_t$. Thus,

$$x_{t-\Delta t}^{\text{DDIM}} \approx x_t + \frac{1}{2} \left(-\frac{\Delta t}{\sigma^2 + t} x_t \right) = x_t \left(1 - \frac{\Delta t}{2(\sigma^2 + t)} \right).$$

(h) Distribution of \hat{X}_0 under DDIM

Starting with $\hat{X}_1 \sim \mathcal{N}(0, 1)$ and applying the DDIM step:

$$\hat{X}_0 = \hat{X}_1 \prod_{k=0}^{T-1} \left(1 - \frac{\Delta t}{2(\sigma^2 + 1 - k\Delta t)} \right).$$

Taking logarithms and approximating $\log(1 - x) \approx -x$:

$$\log \hat{X}_0 \approx \log \hat{X}_1 - \frac{1}{2} \int_0^1 \frac{1}{\sigma^2 + t} dt = \log \hat{X}_1 - \frac{1}{2} \ln \left(\frac{\sigma^2 + 1}{\sigma^2} \right).$$

Thus,

$$\hat{X}_0 \approx \hat{X}_1 \cdot \sqrt{\frac{\sigma^2}{\sigma^2 + 1}} \sim \mathcal{N} \left(0, \frac{\sigma^2}{\sigma^2 + 1} \right) \approx \mathcal{N}(0, \sigma^2).$$

Problem 2: Honey, Where's My Reward Model?

(a) Optimal p for $\min_p D_{\text{KL}}(p\|q)$

The minimum KL divergence is 0, achieved when $p = q$. Thus, $p^* = q$.

(b) Optimal policy form

Maximizing $\mathbb{E}[r_\phi(x, y)] - \beta D_{\text{KL}}[\pi_\theta(y|x)\|\pi_{\text{ref}}(y|x)]$ yields:

$$\pi_{\theta^*}(y|x) = \frac{1}{Z(x)} \pi_{\text{ref}}(y|x) \exp \left(\frac{1}{\beta} r_\phi(x, y) \right),$$

where $Z(x) = \sum_y \pi_{\text{ref}}(y|x) \exp \left(\frac{1}{\beta} r_\phi(x, y) \right)$.

(c) Practical difficulty

The partition function $Z(x)$ sums over the entire output space \mathcal{Y} , which is exponentially large for language models, making computation intractable.

(d) Reward in terms of optimal policy

From the optimal policy expression:

$$r_\phi(x, y) = \beta \log \frac{\pi_{\theta^*}(y|x)}{\pi_{\text{ref}}(y|x)} + \beta \log Z(x).$$

(e) Bradley-Terry model with implicit reward

Substitute $r_\phi(x, y)$ into the Bradley-Terry model:

$$p^*(y_w \succ y_l|x) = \sigma \left(\beta \log \frac{\pi_{\theta^*}(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta \log \frac{\pi_{\theta^*}(y_l|x)}{\pi_{\text{ref}}(y_l|x)} \right).$$

The partition function $Z(x)$ cancels, which is desirable because it avoids intractable computation.

(f) Gradient of DPO loss

Let $\hat{r}_\theta(x, y) = \beta \log \frac{\pi_\theta(y|x)}{\pi_{\text{ref}}(y|x)}$ and $\delta = \hat{r}_\theta(x, y_w) - \hat{r}_\theta(x, y_l)$. The DPO loss is:

$$\mathcal{L}_{\text{DPO}} = -\mathbb{E} [\log \sigma(\delta)].$$

The gradient is:

$$\nabla_\theta \mathcal{L}_{\text{DPO}} = -\mathbb{E} [\sigma(-\delta) (\nabla_\theta \hat{r}_\theta(x, y_w) - \nabla_\theta \hat{r}_\theta(x, y_l))].$$

The weighting term $\sigma(-\delta) = \sigma(\hat{r}_\theta(x, y_l) - \hat{r}_\theta(x, y_w))$ is the probability that the model incorrectly prefers y_l over y_w . The gradient updates more when the model is incorrect (larger $\sigma(-\delta)$).

(g) Plackett-Luce model

Using $r_\phi(x, y) = \beta \log \frac{\pi_{\theta^*}(y|x)}{\pi_{\text{ref}}(y|x)} + \beta \log Z(x)$, we have:

$$\exp(r_\phi(x, y)) = \left(\frac{\pi_{\theta^*}(y|x)}{\pi_{\text{ref}}(y|x)} \right)^\beta Z(x)^\beta.$$

Substituting into the Plackett-Luce probability:

$$p_{\theta^*}(\tau|y_1, \dots, y_K, x) = \prod_{k=1}^K \frac{\left(\frac{\pi_{\theta^*}(y_{\tau(k)}|x)}{\pi_{\text{ref}}(y_{\tau(k)}|x)} \right)^\beta}{\sum_{j=k}^K \left(\frac{\pi_{\theta^*}(y_{\tau(j)}|x)}{\pi_{\text{ref}}(y_{\tau(j)}|x)} \right)^\beta}.$$

The partition function $Z(x)$ cancels out.

Problem 3: Homework Process and Study Group

(a) Sources used

Course notes, textbooks, and online resources.

(b) Collaborators

Worked independently.

(c) Time spent

Approximately 10 hours.

Problem 4: Diffusion Models

(a) Closed-form sampling at time t

Given $q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t I)$, define $\alpha_t = \prod_{i=1}^t (1 - \beta_i)$. By induction,

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t} \mathbf{x}_0, (1 - \alpha_t) I).$$

(b) Reverse conditional distribution

Using Bayes' rule and Gaussian properties:

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t I),$$

where

$$\tilde{\mu}_t = \frac{\sqrt{\alpha_{t-1}} \beta_t}{1 - \alpha_t} \mathbf{x}_0 + \frac{\sqrt{1 - \beta_t} (1 - \alpha_{t-1})}{1 - \alpha_t} \mathbf{x}_t, \quad \tilde{\beta}_t = \frac{1 - \alpha_{t-1}}{1 - \alpha_t} \beta_t.$$