Computer Vision I: Low-Middle Level Vision Homework Exercise #2

(total 10 points)

Due: November 28th 11:59 PM.

Problem 1 (Minimax entropy learning, 3 points).

This question aims to refresh the proof process in minimax entropy learning. Let $p(\mathbf{I})$ be a FRAME model with K histograms matched to the underlying model $f(\mathbf{I})$

$$p(\mathbf{I};\Theta) = \frac{1}{Z(\Theta)} \exp\{-\sum_{i=1}^{K} < \lambda_i, H_i(\mathbf{I}) >\} \tag{1}$$

The parameter $\Theta = (\lambda_1, ..., \lambda_K)$ is learned so that the following constraints are satisfied.

$$E_p[H_i(\mathbf{I})] = E_f[H_i(\mathbf{I})] = h_i, \quad i = 1, 2, ..., K. \tag{2}$$

- **Q1:** Derive the following equation:

$$\frac{\partial \log Z}{\partial \lambda_i} = -E_p[H_i(\mathbf{I})].$$

- **Q2:** Let $\ell(\Theta)$ be the log-likelihood for one observed image $\mathbf{I}^{\text{obs}}$, prove that

$$\begin{aligned} \frac{\partial^2 \ell(\Theta)}{\partial \lambda_i \partial \lambda_j} &= -\frac{\partial^2 \log Z}{\partial \lambda_i \partial \lambda_j} \quad (3) \\ &= -E_p[(H_i(\mathbf{I}) - h_i)(H_j(\mathbf{I}) - h_j)], i, j \in \{1, 2, ..., K\} \quad (4) \end{aligned}$$

  comment: Thus the second derivative of $\ell(\Theta)$ is a negative covariance matrix. So $\ell(\Theta)$ has a single maximum solution.

Now suppose we extract a new feature from the dictionary $F_+ \in \Delta$, and augment the model to

$$p_+(\mathbf{I};\Theta_+) = \frac{1}{Z(\lambda_+)} \exp - \sum_{\alpha=1}^{K} < \lambda_\alpha^*, H_\alpha(\mathbf{I}) > - < \lambda_+, H_+(\mathbf{I}) > \tag{5}$$

The new parameter $\Theta_+ = (\lambda_1^*, ..., \lambda_K^*, \lambda_+)$ is learned to not only satisfy the K constraints specified in equation (2), but also an extra condition:

$$E_{p_+}[H_+(\mathbf{I})] = E_f[H_+(\mathbf{I})] = h_+. \tag{6}$$

Note: To match all the $K+1$ statistical constraints, the existing parameters ($\lambda_\alpha \to \lambda_\alpha^*, i = 1, 2, ..., K$) must be updated when we introduce new features (marginal) because all features are correlated.

- **Q3:** Derive the steps for proving the following theorem

$$KL(f||p) - KL(f||p_+) = KL(p_+||p).$$

Problem 2 (Learning by information projection, 2 points)

Suppose that we are learning the underlying probability model $f(\mathbf{I})$ of image $\mathbf{I}$. We start with an initial probability model, denoted as $q(\mathbf{I})$, and observe that $q(\mathbf{I})$ has a different marginal probability over a macroscopic feature $H_i(\mathbf{I})$:

$$E_q[H_i(\mathbf{I})] \neq E_f[H_i(\mathbf{I})] = h_i,$$

where h_i is estimated from a set of examples sampled from $f(\mathbf{I})$. To improve the current model, we learn a new probability model $p(\mathbf{I})$ so that it reproduces this marginal statistics feature ($p(\mathbf{I})$ may not necessarily replicate all the marginal probabilities that model $q(\mathbf{I})$ has matched previously). We denote the set of models that satisfy this constraint equation by,

$$\Omega_i = \{p : E_p[H_i(\mathbf{I})] = E_f[H_i(\mathbf{I})] = h_i.\}$$

Now, among all the $p(\mathbf{I})$ in Ω_i, we choose one that is closest to $q(\mathbf{I})$ so that it preserves the learning history.

$$p^* = \arg\min_{p \in \Omega_i} KL(p||q) = \arg\min_{p \in \Omega_i} \int p(\mathbf{I}) \log \frac{p(\mathbf{I})}{q(\mathbf{I})} d\mathbf{I}.$$

1. Derive the formula of $p(\mathbf{I})$ by leveraging the Euler-Lagrange equation (Tips: (I) constrained optimization).
2. Prove that $KL(f||q) - KL(f||p) = KL(p||q)$. (Remark: Since $D(p||q) > 0$, p is closer to f than q).
3. Show that this optimization satisfies the maximum entropy principle when $q(\mathbf{I})$ is a uniform distribution.

Problem 3 (Information projection, 3 points)

Considering the feature pursuit in a family of models,

$$p_0(x) \to p_1(x) \to \cdots \to p_K(x) \quad \sim \ f(x).$$

where

$$p_K(x; \Theta_K) = \frac{1}{Z_K} \exp\{-\sum_{i=1}^{K} \lambda_i h_i(x)\}.$$

For simplicity, we treat λ_i as a scalar rather than a vector.

In the minimax entropy process, when we add a new feature statistics $h_K(x)$, we need to update all the parameters $\lambda_i = 1, ..., K$ in the new model $p_K(x; \Theta_K)$ by MLE, so that all the K constraint equations are satisfied,

$$E_{p_K}[h_i(x)] = h_i^{\text{obs}}, \quad i = 1, 2, ..., K.$$

In a different method, we can pursue a series of models in the following way,

$$q_0(x) \to q_1(x) \to \cdots \to q_K(x) \quad \sim \ f(x).$$

with

$$q_K(x) = \frac{1}{z_K} q_{K-1}(x) \exp\{-\beta_K h_K(x)\}.$$

In this model, β_K is decided by the new constraint

$$E_{q_K}[h_K(x)] = E_f[h_K(x)] \approx h_K^{\text{obs}}.$$

In comparison to the previous p-series, the q-series observes the constraints one-by-one, and fixes the previous parameters $\beta_i, i = 1, 2, ..., K-1$ when we learn β_K, i.e.

$$q_K(x;) = \frac{1}{z_1 z_2 \cdots z_K} q_o(x) \exp\{-\sum_{i=1}^{K} \beta_i h_i(x)\}.$$

1. For the q-series, derive the formula for $\frac{\partial \log z_K}{\partial \beta_k}$.
2. Suppose we denote by $Z_K = z_1 z_2 \cdots z_{K-1}$ as the normalizing function for $q_K(x)$, derive $\frac{\partial^2 \log Z_K}{\partial \beta_i \partial \beta_j}$, $\forall i, j \leq K$.
3. Prove that $KL(f||q_K) - KL(f||q_{K+1}) = KL(q_{K+1}||q_K) \geq 0$, and prove the q-series will converge to f

Problem 4 (Typical set, 2 points)

Suppose we toss a coin N times and observe a 0/1 sequence (for head and tail respectively),

$$S_N = (x_1, x_2, ..., x_N), \quad x_i \in \{0, 1\}.$$

S_N is said to be of *type* q (i.e. the frequency of 1 is q in the sequence) with $q = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Let $\Omega(q)$ be the set of all sequences S_N of type q. For simplicity, we discretize q to finite precision.

1. What is the cardinality of $\Omega(q)$ for $q = 0.2$ and $q = 0.5$ respectively? (Suppose we only care about the exponential order or rate).

2. Suppose we know that the underlying probability is $x_i = 1$ (or $x_i = 0$) with probability p (or $1 - p$ respectively), by sampling from this probability N times, what is the probability $p(S_N)$ that we observe a sequence $S_N \in \Omega(q)$? What is the total probability mass $p(\Omega(q))$ for all the sequences in set $\Omega(q)$?

3. In the above question, show that as $N \to \infty$, only sequences from the type p, i.e. set $\Omega(p)$, can be observed.