Problem Set 3

Information and Coding Theory

February 26, 2021 CLAYTON SEITZ

Problem 0.1. *A single dice is rolled and we gain a dollar if the outcome is 2,3,4,5 and lose a dollar if the outcome is 1 or 6. Find the expected gain and the maximum entropy distribution over the possible outcomes of a roll.*

Solution.

Let P be the uniform distribution over the dice universe χ where an outcome of a roll is $x \in \chi$. Furthermore, let $\phi(x)$ be the gain given the outcome of a roll x according the problem definition

$$\phi = \begin{cases} 1 & 2,3,4,5 \\ -1 & 1,6 \end{cases}$$

and $\bar{x} \sim P^n$ be a draw of a sequence of n rolls from the product distribution P^n. We can then calculate the expected gain over n rolls as

$$\begin{aligned} \mathop{\mathbf{E}}_{\bar{x}\sim P^n}[\phi(\bar{x})] &= \sum_n \left(\sum_i \phi(x_n) \cdot p(x_n) \right) \\ &= \sum_n \left(\frac{1}{6} \sum_i \phi(x_n) \right) \\ &= \frac{n}{3} \end{aligned}$$

Now, we would like to find the maximum entropy distribution P^* over χ in the set of distributions Π such that

$$\mathop{\mathbf{E}}_{\bar{x}\sim (P^*)^n}[\phi(\bar{x})] > \frac{n}{3} \tag{1}$$

We can find such a distribution P^* by defining the linear family of distributions that satisfy this constraint on the expected gain

$$\mathcal{L} = \left\{ P : \mathop{\mathbf{E}}_{\bar{x} \sim P^n} [\phi(\bar{x})] = \sum_{x \in \chi} p(x) \cdot \phi(x) > \alpha \right\}$$

We would like to find the distribution P^* such that $P^* = \mathbf{Proj}_{\mathcal{L}}(Q)$ and we now compute this projection by using the Lagrangian

$$\mathbf{\Lambda}(P, \lambda_0, \lambda_1) = D(P||Q) + \lambda_0 \left(\sum p(x) - 1 \right) + \lambda_1 \xi_\alpha(x) \tag{2}$$

where

$$\xi_\alpha = \begin{cases} -x & x < \alpha \\ 0 & x \geq \alpha \end{cases}$$

We find a solution by setting the derivative of this Lagrangian to zero

$$\nabla \Lambda = \log \left(\frac{p^*(x)}{q(x)} \right) + \frac{1}{2 \ln 2} + \lambda_0 + \nabla \xi_\alpha$$

$$\nabla \xi_\alpha = \begin{cases} -\lambda_1 & x < \alpha \\ 0 & x > \alpha \end{cases}$$

Ultimately, we have the solution

$$p^*(x) = q(x) \cdot 2^{\lambda_0 - \lambda_1 \cdot \phi(x)}$$

■

Problem 0.2. *Exponential families and maximum entropy*

Solution.

$$
\begin{aligned}
H(Q) &= -\sum_{x\sim\chi} Q(x) \log \exp\left\{\lambda_0 + \sum_{i\sim[k]} \lambda_i f_i(x)\right\} \\
&= -\frac{1}{\ln 2} \sum_{x\sim\chi} Q(x) \left\{\lambda_0 + \sum_{i\sim[k]} \lambda_i f_i(x)\right\} \\
&= -\frac{1}{\ln 2} \left(\lambda_0 + \sum_{x\sim\chi} Q(x) \left\{\sum_{i\sim[k]} \lambda_i f_i(x)\right\}\right) \\
&= -\frac{1}{\ln 2} \left(\lambda_0 + \sum_{i\sim[k]} \lambda_i \alpha_i\right)
\end{aligned}
$$

Now we will show that the KL-Divergence is the difference of entropies

$$
\begin{aligned}
D(P||Q) &= \sum_{x\sim\chi} p(x) \log \frac{p(x)}{q(x)} \\
&= -\frac{1}{\ln 2} \sum_{x\sim\chi} p(x) \left\{\lambda_0 + \sum_{i\sim[k]} \lambda_i f_i(x)\right\} - H(P) \\
&= -\frac{1}{\ln 2} \left(\lambda_0 + \sum_{i\sim[k]} \lambda_i \alpha_i\right) - H(P) \\
&= H(Q) - H(P)
\end{aligned}
$$

Finally, we can show that Q is the maximum entropy distribution in the family $\mathcal{L}$

$$
D(P||Q) = H(Q) - H(P) \geq 0
$$

which requires that $H(Q) \geq H(P)$.

■

Problem 0.3. *Minimax rates for denoising*

Solution.

This can be shown by using the chain rule for KL-Divergence

$$
\begin{aligned}
D(P(X,Y)||Q(X,Y)) &= D(P(X)||Q(X)) + D(P(Y|X)||Q(Y|X)) \\
&= D(P(Y|X)||Q(Y|X)) \\
&= D(\mathcal{N}(f(x),\sigma^2)||\mathcal{N}(g(x),\sigma^2))
\end{aligned}
$$

which we now compute

$$
\begin{aligned}
D(\mathcal{N}(f(x),\sigma^2)||\mathcal{N}(g(x),\sigma^2)) &= \frac{1}{\ln 2}\int_0^1 \exp\left(-(x-f(x))^2/2\sigma\right) \\
&\cdot \ln\left(\frac{\exp\left(-(x-f(x))^2/2\sigma\right)}{\exp\left(-(x-g(x))^2/2\sigma\right)}\right) dx \\
&= \frac{1}{2\ln 2\cdot\sigma}\int_0^1 \exp\left(-(x-f(x))^2\right) \\
&\cdot \left((x-g(x))^2-(x-f(x))^2\right) dx \\
&= \frac{1}{2\ln 2\cdot\sigma}\int_0^1 |f(x)-g(x)|^2 dx \\
&= \frac{1}{2\ln 2}\cdot ||f(x)-g(x)||_2^2
\end{aligned}
$$

$\blacksquare$