CSE 250B: Homework 3 Solutions

1. *Handwritten digit recognition using a Gaussian generative model.* With a judicious choice of c, a test error rate of about 4% is achievable: close to the performance of nearest neighbor, and much more lightweight.

2. *Entropy examples.*

   (a) $H = (2/3)\log(3/2) + (1/3)\log 3 = 0.9183$

   (b) Six outcomes, equally likely: $H = \log 6 = 2.5850$

   (c) $H = 10$

   (d) $H = 1$

3. *Differential entropy.*

   (a) Let $p(x)$ be the density of a Gaussian with mean μ and variance σ^2. Assuming we are doing logarithms base two, and using $\log z = (\log e)(\ln z)$, we have

$$\begin{aligned}
\int p(x)\log\frac{1}{p(x)}dx &= \int p(x)\log\left(\sigma\sqrt{2\pi}\;e^{(x-\mu)^2/2\sigma^2}\right)dx \\
&= \int p(x)\left(\frac{1}{2}\log(2\pi\sigma^2) + (\log e)\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&= \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log e}{2\sigma^2}\int p(x)(x-\mu)^2 dx \\
&= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log e \;=\; \frac{1}{2}\log(2\pi\sigma^2 e)
\end{aligned}$$

   where we have used the fact that $\int p(x)(x-\mu)^2 dx$ is just the variance of the Gaussian, σ^2.

   (a) Let $p(x)$ the density of the uniform distribution over $[a,b]$. Thus $p(x) = 1/(b-a)$ everywhere in this interval.

$$\int_a^b p(x)\log\frac{1}{p(x)}\,dx \;=\; \int_a^b p(x)\log(b-a)\,dx \;=\; \log(b-a).$$

   This is negative if $b-a<1$: so, differential entropy can be negative!

4. *Distribution of words in an alien language.* The set of possible words, S, consists of all strings over $\{a,b,\ldots,z\}$ of length ≤ 15. For each word $x \in S$, we define the following features:

   - $T_1(x) =$ length of x
   - $T_2(x) = 1(x$ ends with a vowel$)$
   - $T_3(x) = 1($every consonant in x is followed by a vowel$)$
   - $T_4(x) = 1($first letter of x is 'z'$)$

   The observed statistics correspond to the constraints $\mathbb{E}T_1(x) = 4$, $\mathbb{E}T_2(x) = 0.8$, $\mathbb{E}T_3(x) = 0.9$, and $\mathbb{E}T_4(x) = 0.5$. The functional form of the maximum entropy distribution over S is

$$p(x) \;\propto\; \exp(\theta_1 T_1(x) + \theta_2 T_2(x) + \theta_3 T_3(x) + \theta_4 T_4(x)).$$

5. Define $S = [0,\infty)$, with base measure $\pi \equiv 1$. Let $T(x) = x$. Then the exponential family generated by (S,π,T) consists of distributions $\{p_\theta\}$ where $\theta \in \mathbb{R}$ and

$$p_\theta(x) = \frac{1}{Z_\theta}e^{\theta x}.$$

The normalizer is given by

$$Z_\theta = \int_0^\infty e^{\theta x}\, dx = \left[\frac{e^{\theta x}}{\theta}\right]_0^\infty = \begin{cases} \infty & \text{if } \theta \geq 0 \\ -1/\theta & \text{if } \theta < 0 \end{cases}$$

Therefore the natural parameter space is $\Theta = \mathbb{R}^-$, and we have

$$p_\theta(x) = -\theta e^{\theta x}.$$

These are exactly distributions of the specified form, with $\lambda = -\theta$.

6. $(1/\sqrt{14}, 2/\sqrt{14}, 3/\sqrt{14})$
7. $(-1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$
8. $x \cdot x = 25 \Leftrightarrow \|x\| = 5$. All points of length 5: a sphere, centered at the origin, of radius 5.
9. $f(x) = 2x_1 - x_2 + 6x_3 = w \cdot x$ for $w = (2, -1, 6)$.
10. A is 10×30 and B is 30×20
11. (a) X is $n \times d$
    (b) XX^T is $n \times n$
    (c) $(XX^T)_{ij} = x^{(i)} \cdot x^{(j)}$
12. $((x^Tx)(x^Tx)(x^Tx)) = (\|x\|^2)^3 = 10^6$
13. $x^Tx = \|x\|^2 = 35$ and

$$xx^T = \begin{bmatrix} 1 & 3 & 5 \\ 3 & 9 & 15 \\ 5 & 15 & 25 \end{bmatrix}$$

14. The angle θ between x and y satisfies $\cos\theta = x^Ty/\|x\|\|y\| = 1/2$, so θ is 60 degrees.
15.

$$M = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 0 & 0 \\ -2 & 0 & 6 \end{bmatrix}$$

16. *Symmetric Matrices*
    (a) $(AA^T)^T = (A^T)^TA^T = AA^T$, Thus AA^T is symmetric.
    (b) $(A^TA)^T = A^T(A^T)^T = A^TA$, Thus A^TA is symmetric.
    (c) $(A + A^T)^T = (A^T + A) = (A + A^T)$, Thus $(A + A^T)$ is symmetric
    (d) $(A - A^T)^T = (A^T - A) \neq (A - A^T)$, Thus $(A - A^T)$ need not be symmetric
17. (a) $|A| = 8! = 40320$
    (b) $A^{-1} = \text{diag}(1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8)$
18. *Orthonormal matrices*
    (a) UU^T is the identity matrix
    (b) $U^{-1} = U^T$
19. Since A is singular matrix, $|A| = 0 \implies z - 6 = 0 \implies z = 6$