Math Review Solutions.
Information Theory Problem Set 1.

1. Let

$$J(p_1, \ldots, p_n, \lambda) = \sum_{k=1}^{n} p_k log(p_k) - \lambda \left(\sum_{k=1}^{n} p_k - 1 \right)$$

where λ is a Lagrange multiplier. Differentiating J and setting the derivative to 0 yields

$$\frac{\partial J}{\partial p_m} = p_m \frac{1}{p_m} + log(p_m) - \lambda = 0$$

$$\implies \lambda - 1 = log(p_m) \implies 2^{\lambda-1} = p_m$$

Applying the constraint

$$\sum_{m=1}^{n} 2^{\lambda-1} = \sum_{m=1}^{n} p_m \implies n2^{\lambda-1} = 1.$$

Finally, it follows from $2^{\lambda-1} = p_m$ that $np_m = 1$ so $p_m = \frac{1}{n}$.

2. Mutual Information is given by $\mathcal{M}(\mathbf{p}, \mathbf{q}) = \sum_m \sum_n f(p_m, q_n) log\left(\frac{f(p_m, q_n)}{p_m q_n}\right)$. It can be calculated as follows:

```
MI1 = 0;
for m = 1:3;
    pm = p(m);
    for n = 1:3;
        qn = q1(n);
        MI1= MI1+pq1(m,n)*log2(pq1(m,n)/(pm*qn));
    end;
end;
```

which gives the answers $\mathcal{M}(\mathbf{p}, \mathbf{q}_1)$ = 0.0865 and $\mathcal{M}(\mathbf{p}, \mathbf{q}_2)$ = 0. If one uses the natural logarithm (base *e*) instead of base 2, the answers are $\mathcal{M}(\mathbf{p}, \mathbf{q}_1)$ = 0.0599 and $\mathcal{M}(\mathbf{p}, \mathbf{q}_2)$ = 0.

3. (a) An example of a convex function is shown in C.8

   (b) To develop intuition, observe the plots of entropy for $\mathbf{p} = (p, (1-p))$ and $\mathbf{p} = (p_1, p_2, 1-(p_1+p_2))$ as shown in C.9:

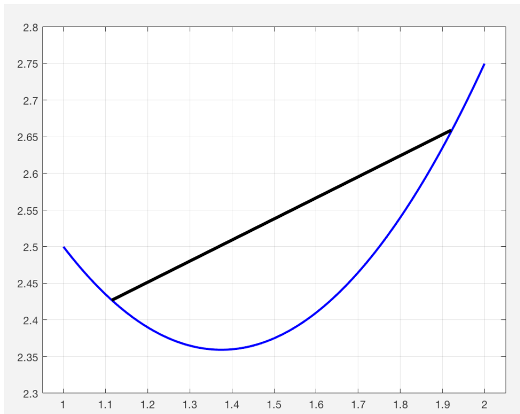

Figure C.8 An example of a convex function.

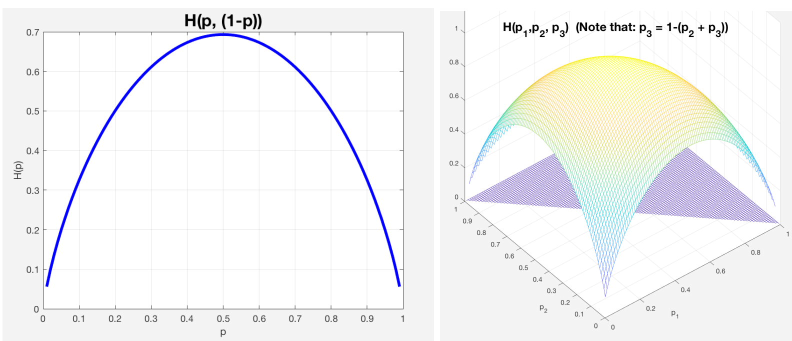

Figure C.9 Plots of Entropy for 2 and 3 variable pdfs.

The plots suggest that entropy is concave although it is not a proof. One argument is: Consider a single term of entropy, $f(p_m) = -p_m log(p_m)$. The derivative of f is $\frac{df}{dp_m} = -(1 + log(p_m))$. The second derivative is $\frac{d^2}{dp_m^2} = -\frac{1}{p_m} < 0$ assuming $p_m \neq 0$. Hence, $\frac{df}{dp_m}$ is strictly decreasing so f is concave. Suppose f and g are concave functions of a single, real-valued variable with domain $A \subset \mathbb{R}$. Then

$$(f+g)(ax + (1-a)y) =$$

$$f(ax + (1-a)y) + g(ax + (1-a)y) \geq$$

$$af(x) + (1-a)f(y) + ag(x) + (1-a)g(y) =$$

$$a(f+g)(x) + (1-a)(f+g)(y)$$

so $f+g$ is concave. Therefore, entropy is concave. Do you believe this argument? How about the argument in part (c)? We'll discuss in class.

(c) Mutual Information can be written as follows

$$\mathcal{M} = \sum_x \sum_y p(x,y)\, log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) = \sum_x \sum_y p(y|x)\, p(x)\, log\left(\frac{p(y|x)}{p(y)}\right) =$$

$$= \sum_x p(x) \sum_y p(y|x)\, log\left(\frac{p(y|x)}{p(y)}\right) = \sum_x p(x)\, \mathrm{KL}\left(p(y|x)\,, p(y)\right) =$$

$$= \mathbb{E}_{p(x)}\left[\mathrm{KL}\left(p(y|x)\,, p(y)\right)\right].$$

The last expression shows that $\mathcal{M}$ is a linear function of $p(x)$. Therefore it is both convex and concave as a function of $p(x)$.The derivative of $\mathcal{M}$ with respect to $p(y|x)$ for a specific $y = y_d$ and $x = x_d$ is $\frac{d\mathcal{M}}{dp(y_d|x_d)} = p(x_d)\left[1 + log\left(p(y_d|x_d)\right) - log(y_d)\right]$ and the second derivative is $\frac{p(x_d)}{p(y_d|x_d)} \geq 0$. Therefore, $\mathcal{M}$ is convex with respect to $p(y|x)$ for any y and x.

4. (a) Note that $H(Y|X=x) = -\sum_y p(y|x)\, log\left(p(y|x)\right)$. Take $p(x) = p_x$ for $x = 1,2,3$ and $p(y) = p_y$ for $y = 1,2,3$. Then $p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(x,y)}{p_x}$. For $\mathbf{p}$ and $\mathbf{q}$, the latter expression is

$$p(Y|X) = \begin{bmatrix} 0.30 & 0.40 & 0.30 \\ 0.02 & 0.68 & 0.30 \\ 0.10 & 0.60 & 0.30 \end{bmatrix}$$

where the element in row r and column c represents $\frac{p(r,c)}{p_r}$. Then

$$\mathbf{h} = \left(H(Y|X=1)\,, H(Y|X=2)\,, H(Y|X=3)\right)^t = (1.57, 1.01, 1.30)^t$$

and

$$H(Y|X) = \mathbb{E}_{p(x)}\left[H(Y|X=x)\right] = \mathbf{p}^t\mathbf{h} = 1.21.$$

(b) Applying algebra and using the fact the $\sum_y p(y|x) = 1$ yields $H(X,Y) = H(Y|X) + H(X)$. Since entropy is non-negative, it must be true that $H(X,Y) \geq H(Y|X)$.

5. it is easy to construct a joint pdf with Mutual Information 0. Take 2 pdfs of one variable, $p_1(x)$ and $q_1(y)$ and compute all possible products, $p(x,y) = p_1(x)\, q_1(y)$. To create one with Mutual Information non-zero, modify the one with zero by adding and subtracting values while maintaining the stochastic constraints. For example, starting with $p(x,y)$, let d be any number satisfying $0 < d < min\left(p_1(x)\, q_1(y)\right)$ and take

$$
\begin{aligned}
q(x_1, y_1) &= p_1(x_1)\, q_1(y_1) + d \\
q(x_2, y_1) &= p_1(x_2)\, q_1(y_1) - d \\
q(x_1, y_2) &= p_1(x_1)\, q_1(y_2) - d \\
q(x_2, y_2) &= p_1(x_2)\, q_1(y_2) + d
\end{aligned}
$$

and $q(x, y) = p(x, y)$ otherwise.

6. See the solution to problem 3.