Solution 1:

1. $\mathrm{H}(P,Q) = \mathrm{H}(P \mid Q) + \mathrm{H}(Q) = \mathrm{H}(Q \mid P) + \mathrm{H}(P)$ We can use direct proof to show that the above statement is true.
   Assuming that
   $$\mathrm{H}(P,Q) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(P=i,Q=j)logP(P=i,Q=j) \text{ and}$$
   $$\mathrm{H}(P) = -\sum_{i=1}^{n} p_i logp_i \text{ and}$$
   $$\mathrm{H}(P \mid Q) = \sum_{j=1}^{m} q_j \,\mathrm{H}(P \mid Q=j)$$
   are true, and we can therefore state that
   $$\mathrm{H}(Q) = -\sum_{j=1}^{m} q_i logq_i \text{ and}$$
   $$\mathrm{H}(Q \mid P) = \sum_{i=1}^{n} p_i \,\mathrm{H}(Q \mid P=i)$$

   Lets first prove that the joint entropy of some set P, Q of discrete probabilities bound by n and m equals the the conditional entropy of P given Q plus that entropy of Q. Equivalently, it could be said that once Q is known, we only need $\mathrm{H}(P,Q) - \mathrm{H}(Q)$ bits to describe the whole system. Let's start by solving:

Proof.

$$
\begin{aligned}
\mathrm{H}(P,Q) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(i,j) logp(i,j) && \text{as per above definition}\\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(i,j) logp(i)p(j,i) && \text{translate joint entropy}\\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(i,j) logp(j,i) - \sum_{i=1}^{n}\sum_{j=1}^{m} p(i,j) logp(i) && \text{Separate using the log rule}\\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(i,j) logp(j,i) - \sum_{i=1}^{n} p(i) logp(i) && \text{The sum of probabilities equals 1}\\
&= \mathrm{H}(Q \mid P) + \mathrm{H}(P)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{H}(Q,P) &= -\sum_{j\in Q, i\in P} p(j,i) logp(j,i)\\
&\dots\\
&= -\sum_{j\in Q} p(j,i) logp(i,j) - \sum_{j\in Q} p(j) logp(j)\\
&= \mathrm{H}(P \mid Q) + \mathrm{H}(Q)
\end{aligned}
$$

intuitively,

$$
\begin{aligned}
\mathrm{H}(P,Q) &= \mathrm{H}(Q,P)\\
\therefore \mathrm{H}(P,Q) &= \mathrm{H}(P \mid Q) + \mathrm{H}(Q) = \mathrm{H}(Q \mid P) + \mathrm{H}(P)
\end{aligned}
$$

□

2. If P and Q are independent, then $\mathrm{H}(P \mid Q) = \mathrm{H}(P)$
   If P and Q are independent sets, there is no overlap and the probability of one would not dictate the other, it is therefore intuitive that the set of probabilities would not be impacted by another set by simply rearranging the above equation we can prove this.

   Proof.

$$
\begin{aligned}
\mathrm{H}(P \mid Q) &= \mathrm{H}(P) * \mathrm{H}(Q) && \text{given independence}\\
&= -\sum_{i=1}^{n} p(i) log(i) * -\sum_{j=1}^{m} p(j) log(j)\\
&= \sum_{i=1}^{n} p(i) log(i) * 1 && \text{as the probabilities of a set sum to 1}\\
&= \mathrm{H}(P)
\end{aligned}
$$

$\square$

3. $\max\{\mathrm{H}(P), \mathrm{H}(Q)\} \leq \mathrm{H}(P, Q) \leq \mathrm{H}(P) + \mathrm{H}(Q)$ *Hint*: You can assume that the inequality $\log x \leq x - 1$ holds for all $x \in \mathbb{R}$, with equality if and only if $x = 1$.

   Proof. The statement argues that the maximum of P or Qs entropy will always be less than or equal to the

   $\square$

Solution 5:

Describe (and include the plots) the five learning curves by answering these questions for 1.–5. (if applicable):

- describe the models performance
- is there overfitting?
- how does this complexity compare to the baseline 1.?
- how could you improve this model?

Finally, experiment with your own values of d and n until you find what you think is a good model (and better than (a)–(f)). Explain in what ways you think it is good, and explain how think its performance relates to your choice of d and n as well as the data set.