4. a) $P(a,b|c) = \dfrac{P(a,b,c)}{P(c)}$

$$P(a|b,c)\cdot P(b|c) = \frac{P(a,b,c)}{P(b,c)} \cdot \frac{P(b,c)}{P(c)} = \frac{P(a,b,c)}{P(c)}$$

So $P(a,b|c) = P(a|b,c)\cdot p(b|c)$

b) $\dfrac{P(a|b,c)\cdot p(b,c)}{P(a|c)} = \dfrac{P(a,b|c)}{P(a|c)} = \dfrac{P(b,a|c)}{P(a|c)} = P(b|a,c)$

c) When ordinary least squares overfits, $\vec{w}$ contains elements with large magnitude.

d) Testing loss

e) Use 70% of the data for training, and 30% for validation. Then choose the feature with lowest validation loss.