1 Summary: Probability

• if Ω is a set of outcome/events and $A \subset \Omega$ then uniform probability $Pr[x \in A] = |A|/|\Omega|$. There are 2 options here for each outcome, either $x \in A$ or $x \notin A$

• random variables X, Y, X each partition the space Ω by a certain criteria (for example X = object color, Y= object price, Z = object shape). They are called "random" because any object pulled at random can have any of the values of X as color, any value of Y as price, and any Z as shape. For short notation we say $Pr(X, Y) = Pr(x = X, y = Y)$ is the probability that a random object has color x and price y

• Bayes Theorem $Pr(Y|X) * P(X) = Pr(X, Y) = Pr(X|Y) * P(Y)$ which means probability to have a particular $(X = red, Y = 100)$ object is the probability to have $(X = red)$ times probability to have $(Y = 100,$ given that $X = red)$ or vice versa. The equality is the same as saying $Pr(Y|X) = \frac{Pr(X|Y)*P(Y)}{P(X)}$

• marginalization of variable Y over variable
X: $Pr(Y) = \sum_x Pr(x = X) * Pr(X, Y) = \sum_x Pr(x = X) * Pr(Y|X)$.
If X is binary with only two possible values $(X; \bar{X})$ (for example "pass" vs "fail") then we have
$Pr(Y) = Pr(X) * Pr(Y|X) + Pr(\bar{X}) * Pr(Y|\bar{X})$

• with marginalization of Y over binary random variable X we can write Bayes as
$Pr(X|Y) = \frac{Pr(Y|X)*P(X)}{P(Y)} = \frac{Pr(Y|X)*P(X)}{Pr(X)*Pr(Y|X)+Pr(\bar{X})*Pr(Y|\bar{X})}$

• Independent variables X, Y means $P(Y|X) = P(Y)$ which is to say Y does not depend on X (price does not depend on color). From Bayes this means also $P(X|Y) = P(X)$ and $P(X, Y) = P(X) * P(Y)$

• expected value (or mean) for a numeric-value random variable is the average of values weighted by probabilities:
$E[X] = \sum_x x * Pr(x = X) = \sum_x x * Pr(x)$

• expectation ALWAYS distributes over sum, even if variables are not independent. But they have to have numeric values
$E[X + Y + Z] = E[X] + E[Y] + E[Z]$

• variance = avg distance-to-mean^2 weighted by probabilities
$\text{var}[X] = \sum_x (x - mean)^2 * Pr(x = X) = E[(X - E[X])^2] =$
$E[X^2 + E^2[X] - 2XE[X]] = E[X^2] + 2E^2[X] - 2E^2[X] = E[X^2] - E^2[X]$

• variance distributes over sum ONLY IF X, Y are indenpendent
X, Y independent $\Rightarrow \text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$

• entropy = randomness of X (the more random, the higher the entropy)
$H[X] = \sum_x Pr(x) \log(\frac{1}{Pr(x)})$

• Markov Chains