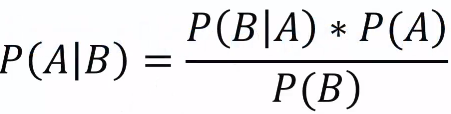
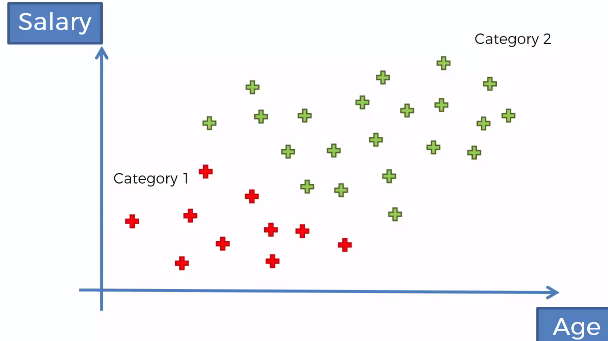
***Bayes’ Theorem***

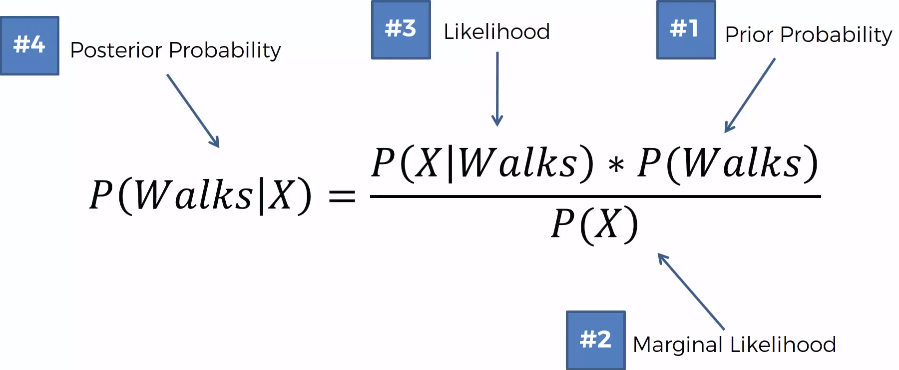


* Given info:
* Mach1 = 30 wrenches/hr 🡪 P(1) = .6
* Mach2 = 20 wrenches/hr 🡪 P(2) = .4
* 1% of parts produced are defective 🡪 P(D) = .1
* 50% of defective parts come from Mach1 and 50% from Mach2 🡪 P(D|2) = .5, P(D|2) = .5
* What is probability a part is defective is from Mach2?
* P(D|2) = P(2|D) \* P(D) / P(2) = (.5 \* .01) / .4 = **.125**
* So if machine 2 produces 10k parts, 125 will be defective
* Ex: 1k wrenches 🡪 400 from M2 🡪 1% have defect = 10, half are from M2 = 5, so % of defects of M2 = 5/400 = .125

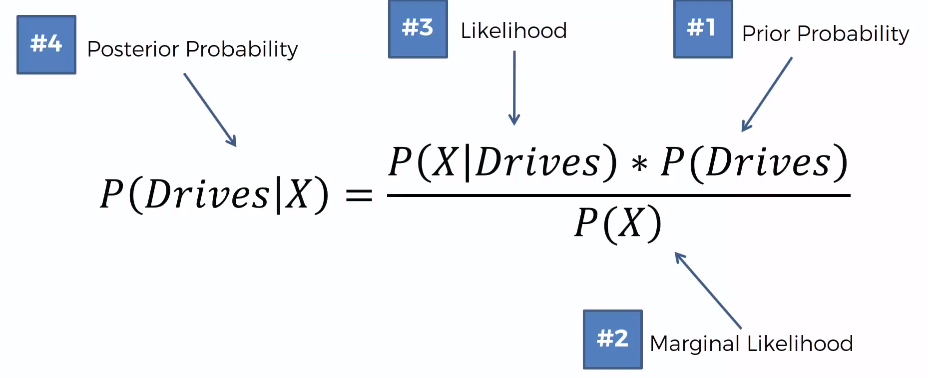
**Naïve Bayes**

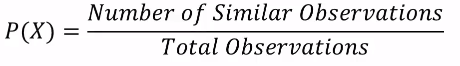


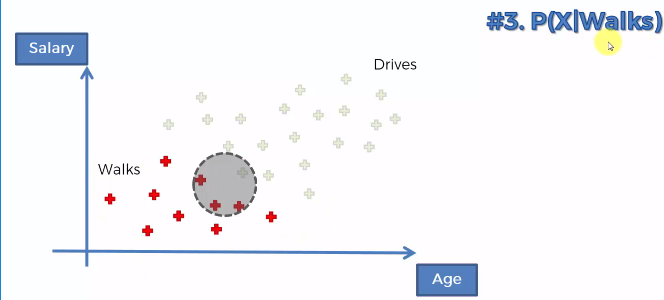
* We will take Bayes Theorem approach *twice:*

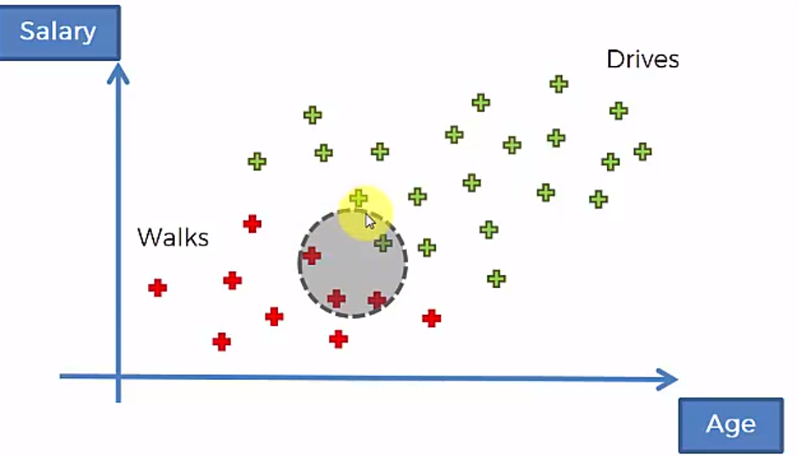


* What is the probability a person walks given their age or given their salary (feature X)
* Then do same for other class



* Prior1 🡺 P(Walks) = # of “walkers” / total observations = 10 / 30
* Marginal Likelihood1 🡺 Create a circle around new DP and deem DP’s near it to be similar and have similar features, so probability of a new DP being class X = # of similar observations / total observations:  = 4/30
* Likelihood1 🡺 Same circle as above, find probability of a random DP from the data set being similar to the new DP



* Given we’re only working w/ “walkers”, what is the probability a random DP from the “walkers” exhibits similar features to the new DP = # of similar observations among those who walk / total walkers = 3/10
* So posterior = 3/10 \* 10/30 / 4/30 = .3 \* .33 / (2/15) = .75
* For drivers, we have (1/20 \* 20/30) / (4/30) = .25 (P(X) is the same)
* So it’s more likely that this new DP walks to work
* Why “Naïve”?
* b/c it requires some independence assumptions, from the Bayes’ Theorem
* Since these are oftentimes not correct, we says it’s naïve to assume our assumptions our correct
* Ex: Naïve Bayes assume age and salary are independent, which usually is not the case IRL
* We can’t *really* apply Bayes’ Theorem since there’s some correlation, but we call the algorithm naïve b/c oftentimes its applied to variable/features that are NOT independent
* Remember, P(X) = probability a randomly selected DP from the dataset will exhibit feature values similar to the new DP being added
*  What is the likelihood of landing in the grey circle when we throw a new DP into the dataset
* We could remove this from out comparison calculations since it’s in both sets, but *only if comparing + NOT interested in actual values*
* 