So the probblem that we want to solve now is basically finding out

$$P(X = x/y = a_i)$$

,

That in itself is not an easy problem to solve,
So X = x, what does this mean? so our input is 2 dimensional, it has few features, lets assume n features.
This actually means the probablity $P((f_1, f_2, f_3 ... f_n) = (x_1, x_2, x_3 ... x_n)/y = ai)$. Now lets come to the part where this classifier is callled $Naive\ Bayes$ Classifier.
It makes a very strong assumption(Naive Assumption) that lets assume that all the features are independent of each other.
And for Independent events, for eg what is the probablity that A happens and B happens $P(A \cap B)$ if A and B are independent of each other, this probablity will be $P(A) * P(B)$.
Thats like saying what is the probablity that I tossed a coin, get a head, and the next time I tossed a coin I get tails.
Both these events are independent of each other.
The probablity that first time I get a head and the second time I get a tail, what you can do is given these two events are independent of each other is $P(1st = Head) * P(2nd = tail)$ = 1/2 *1/2 which is 0.25 .*
Similarly here if we assume that all the features are independent of each other, the probablity $P((f_1, f_2, f_3 ... f_n)$ *is equal to $$ P(f{1} = x{1}/ y = a_{i})* P(f{2} = x{2}/ y = a*{i}) * P(f*{3} = x{3}*/ y = a*{i}) ...... P(f*{n}* = x{n}/ y = a*{i})*

$$. Soweareassumingallthefeaturesareindependentofeachotherandgiventhefacttheyareindependentofeachother, wecanjustmultiplytheseprobablitieswithoutworryingaboutifonehappeningtellsmemoreabouttheotherhashappenedornot. So\$Naive\$\$Bayes\$iscalled\$Naive\$becauseinreality, thatsnotgoingtobethecase. Thesefeaturesaregoingtobedependuponeachother, mostofdatasetwonthavepropertyofIndependenceofthefeaturesthatwehave. ThatsthereasonwecallitNaiveBayes . So,$$

P(X = x / y = a{i}) = \prod*{j = 1}^n P(x^{j} = x^{j} / y = a*{i})

$$Sofor\$P(X = x/y = a_i) * P(y = a_i)\$,$$

P(X = x / y = a*{i}) *P(y=a*{i}) = \prod*{j = 1}^n P(x^{j} = x^{j} / y = a*{i})*P(y=a*{i})$$. So now we dont have to worry about $P(X = x / y = a*{i})$, we have to worry about Individual features to find out the final no that we want to calculate instead of all features combined.
We will find out abt these individual features next.