- Categorical data: ordinal, nominal and binary.
- Generative classifiers. Steps in modeling: model selection; density estimation of the data; classification of new data.
- **Bayesian Theorem**. $p(c|x) = \frac{p(c,x)}{p(x)} = \frac{p(x|c)p(c)}{p(x)}$. c: the model to be inferred. x: the observations. $p(x|c)$: the likelihood. $p(c)$ the prior. $p(c|x)$ the posterior. $p(x)$ the evidence.
- Something more about Bayes. $p(c|x)$ means the probability of c given x. Take a concrete example. $p(L|W) = 0.75$ (from Wikipedia)—the probability of a woman with long hair is 75%. Or rather, the probability of event L given event W is 0.75.
- Even something more. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. $P(A)$ the prior, the initial degree of belief in A. $P(A|B)$ the posterior, the degree of belief having accounted for B.
- **Bayes decision rule**. $p(x)$ is independent of the class and $p(c)$ is frequently assumed to be the *same* for all classes.
- Standard *k-NN*. When we use big k, then we could have a better noise resistance. At the same time we have worse resolution.
- Kernel density estimation. When we increase the number of kernels, better smoothness is obtained. But we will have a much higher dimension. (You know what this implies.)
- Calculate the joint probability distribution is normally very difficult, but Naive Bayesian method solves this by assuming that class conditional distributions are all independent. $p(c_i|x) = \frac{p(x_1,x_2,\ldots,x_n|c_i)p(c_i)}{p(x)} = \frac{p(x_1|c_i)p(x_2|c_i)\ldots p(x_n|c_i)p(c_i)}{p(x)}$.
- Kind note about Naive Bayesian classifier. The evidence $p(x)$ is class-independent, which means it has nothing to do with, the variable. Aka, the stuff that we *should* really look into.
- **Multivariate Bernoulli** model. For this model, one weakness is that whether a word (in Anti-Spamming case) appears 1 or 100 times, the final probability representation is the same. (Let's do it in Chinese. It's NOT scientific.)
- **Multinomial** event model. For this model, each "word" (in spamming example) is an event. One word appears, then it's probability is calculated. Appears many times, then times as many.