Naive Bayes Quiz

1. Naive Bayes adopts a generative approach to classification.

2. A generative classifier
   - is probabilistic, but not all probabilistic classifiers are generative.
   - constructs a model of each class

3. A Bayesian classifier that chooses between several classes makes its choice based on:
   - the probability of each class given the att. values

4. and 5. The Bayesian formulation of $P(y|x)$:

$$P(y|x) = \frac{P(x|y)\,P(y)}{\sum_{y'} P(x|y')P(y')}$$

- $P(x|y)$: how likely it is to see item x if the class is y. (class-conditional probability)
- $P(y)$: prior class probability: how common class y is
- $\sum_{y'} P(x|y')P(y') = P(x)$: normalizing factor: how common item x is

6. If we present a Bayesian classifier with an instance that is very unlike any instance it was trained on, the value of the normalizer ($\sum_{y'} P(x|y')P(y')$) will be small.

7. Naive Bayes ~ Bayesian classifier with a particular assumption
   the attributes are conditionally independent given the class
   → assume attributes are independent ↛ naive

8. To create a classifier that uses Bayes rule, we have to assume something about the probability distribution of each class in the population.
   → assume uniform distribution

9. The Gaussian Naive Bayes model assumes:
   - For a given class, it is reasonable to model each of the atts w/ its own Gaussian.
   - Each class is modelled separately: it is a class-conditional model
   - Given a class, each att. is conditionally independent.
     → each has its own univariate Gaussian

10. Decision boundary in probabilistic classifier represents the points where classes are equally probable.

11. If atts have discrete values, NB calculates probability of a particular att. value given a class $P(x|y)$ by counting how many instances of the class have each attribute value.

12. Missing att. values for dp at training time - ignore, don't add anything to count.
    Att. w/ zero-count in a class - add a small amount to each count.
    * Zipf's law: half the word occur once in a corpus of documents means zero-counts are frequent in that domain.

* How is Naive Bayes a linear classifier?

$\sigma(a) = \frac{1}{1+\exp(a)}$

$$p(c=1|x) = \frac{p(x|c=1)p(c=1)}{p(x|c=1)p(c=1) + p(x|c=0)p(c=0)}$$

$$= \frac{1}{1 + \frac{p(x|c=0)p(c=0)}{p(x|c=1)p(c=1)}}$$

then to turn this into a 'squashing' function

$$= \frac{1}{1 + \exp\left(-\log\left[\frac{p(x|c=1)\,p(c=1)}{p(x|c=0)\,p(c=0)}\right]\right)}$$

$$= \sigma\left(\sum_i \log \frac{p(x_i|c=1)}{p(x_i|c=0)} + \log \frac{p(c=1)}{p(c=0)}\right)$$