More generative modeling

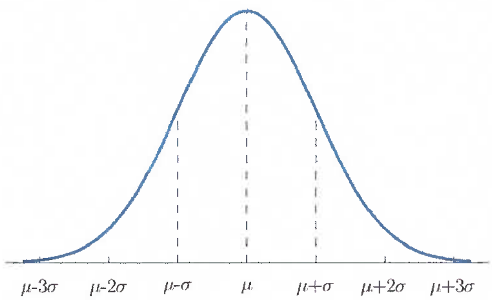
Topics we'll cover

1. Beyond Gaussians
2. A variety of univariate distributions
3. Moving to higher dimension

Classification with generative models

- Fit a **distribution** to each class separately
- Use Bayes' rule to classify new data

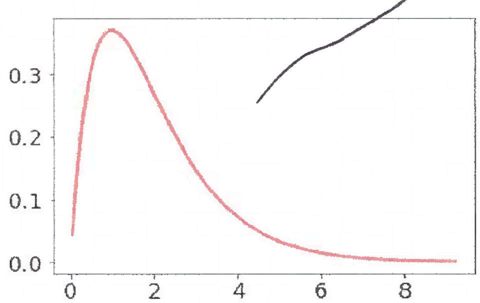
What distribution to use? Are Gaussians enough?



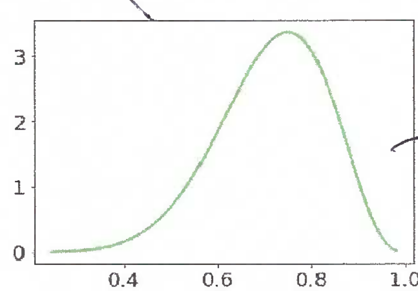
real number including gaussian

Exponential families of distributions

— have a particular form

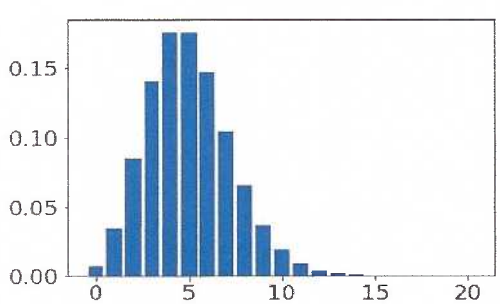


GAMMA



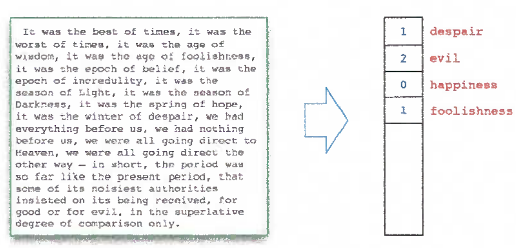
BETA

distributions on an interval



POISSON

data are integer



CATEGORICAL

← not numeric

Multivariate distributions

We've described a variety of distributions for **one-dimensional** data.
What about higher dimensions?

1. **Naive Bayes**: Treat coordinates as independent. Independent
   For $x = (x_1, \ldots, x_d)$, fit separate models $\Pr_i$ to each x_i, and assume
   $$\Pr(x_1, \ldots, x_d) = \Pr_1(x_1)\Pr_2(x_2)\cdots\Pr_d(x_d).$$
   This assumption is typically inaccurate. → But quite effective in practice
2. **Multivariate Gaussian**.
   Model correlations between features: we've seen this in detail. Dependent - Correlates
3. **Graphical models**.
   Arbitrary dependencies between coordinates. probability distributions over many coordinates