



menu

› Course Shortcuts

› Student Lounge

› Q&A

› Estimating Distributions from Data

› Classification with the Naive Bayes Algorithm

› Building a Baby Name Classifying System

📖 Module Introduction: Building a Baby Name Classifying System

📖 Build a Baby Name Classifier

📖 Determine Class Prior Probability

↔️ Class Prior Probability

📖 Determine Conditional Probability

↔️ Conditional Probability

📖 Create Feature Vectors from Data

↔️ Feature Extraction and Hashing

📖 **Baby Name Classifier Walkthrough**

🔗 [Build a Baby Name Classification System](#)

💬 [Improving Focus on Features](#)

📖 [Module Wrap-up: Building a Baby Name Classifying System](#)

📖 [Thank You and Farewell](#)

🔖 [Course Exit Survey](#)

🌐 [Stay Connected](#)

› Course Resources

Estimating Probability Distributions ›

Having seen the different parts of the Baby Name Classifier separately, you might be wondering how everything comes together to create a classifier that you can use for prediction. We will use the hashing, NaiveBayesPY, and NaiveBayesPXY functions and walk through example names. You will then implement the classifier in the final assignment.

Data

Let's start with checking what our data looks like. You will be given two files with train data `girls.train` and `boys.train` containing a baby name in each line. The assignment will also load two files with test data in the same format. For example, the `girls.train` file might look as follows (actual names can vary in the assignment):

Addisyn
Danika
Emilee
Aurora
Julianna
Sophia
Kaylyn
Litzzy
Hadassah

Training

In the assignment, we will use +1 label for boys and -1 label for girls. This is the order in which we will create the classifier's constituent functions:

- Hash each name to a d -dimensional feature vector.
- Compute NaiveBayesPY values $P(Y = 1)$ and $P(Y = -1)$.
- Compute NaiveBayesPXY values $P([\mathbf{x}]_{\alpha} | Y = 1)$ and $P([\mathbf{x}]_{\alpha} | Y = -1)$ for all features $1 \leq \alpha \leq d$.

Now we have trained our classifier. For prediction, we will compute the log likelihoods $\log P(\mathbf{x}_{test} | Y = 1)$ and $\log P(\mathbf{x}_{test} | Y = -1)$. This will then help us compute the predictions $P(Y = 1 | \mathbf{x}_{test})$ and $P(Y = -1 | \mathbf{x}_{test})$ using the Bayes' theorem.

Hashing

For this example, we will use 3 features:

- $[\mathbf{x}]_1$ is 1 if first letter is 'a'
- $[\mathbf{x}]_2$ is 1 if first letter is a vowel (one of 'aeiouy')
- $[\mathbf{x}]_3$ is 1 if last letter is a vowel (one of 'aeiouy')

The feature vectors for these 4 names will be:

	$[\mathbf{x}]_1$	$[\mathbf{x}]_2$	$[\mathbf{x}]_3$
Alicia	1	1	1
Eve	0	1	1
Aiden	1	1	0
Fred	0	0	0

NaiveBayesPY

For simplicity, let's assume we are given 600 girls and boys names each. Therefore, $P(Y = 1) = P(Y = -1) = 0.5$.

NaiveBayesPXY

Using the train data, we can calculate what the probability is of a name starting with 'a', given that it is a boy's name. This is exactly $P([\mathbf{x}]_1 = 1 | Y = 1)$. $P([\mathbf{x}]_1 = 0 | Y = 1)$ is just $1 - P([\mathbf{x}]_1 = 1 | Y = 1)$ as probabilities of different values of the same feature given a label sums to 1. **Empirically, we just find the fraction of boys' names that start with 'a'**.

Similarly we can calculate the conditional probabilities of each feature given the name is a girl's.

We have calculated these empirical conditional probabilities from the train data for you:

	$P([\mathbf{x}]_1 = 1 Y = y)$	$P([\mathbf{x}]_2 = 1 Y = y)$	$P([\mathbf{x}]_3 = 1 Y = y)$
$y = 1$ (boy)	0.09	0.17	0.30
$y = -1$ (girl)	0.16	0.24	0.73

Testing

To predict $P(Y | \mathbf{x})$ — whether the following names are girls' or boys' — we find the log probabilities of each name being a girl's or a boy's and compare. If a name has a larger log probability for a girl (thus a larger probability for a girl), the classifier will predict -1 or "girl"; vice versa for boy.

For any label y , $\log P(Y = y | \mathbf{x}) \propto \log P(Y = y) + \sum_{\alpha=1}^d \log P([\mathbf{x}]_{\alpha} | Y = y)$. Therefore, to predict a label y for a test point \mathbf{x}_{test} , we need to find $\log P(Y = y | \mathbf{x}_{test})$ for different labels y and decide which log likelihood is the largest.

We calculated $P(Y = y)$ in the NaiveBayesPY section, and $P([\mathbf{x}]_{\alpha} | Y = y)$ in the NaiveBayesPXY section.

If a test name's feature $[\mathbf{x}]_1$ is 1, then $P([\mathbf{x}]_1 | Y = 1) = P([\mathbf{x}]_1 = 1 | Y = 1)$. If it is 0, then $P([\mathbf{x}]_1 | Y = 1) = P([\mathbf{x}]_1 = 0 | Y = 1)$. Informally, $P([\mathbf{x}]_1 | Y = 1)$ is the probability of the first feature value of the test name being whatever value it is, assuming it is a boy's name. We know probabilities for all possible values from the training data (computed in NaiveBayesPXY), and we just select the correct value.

Implementation tip: As all our features are binary, there is a clever way for computing $P([\mathbf{x}]_1 | Y = 1)$:

$$P([\mathbf{x}]_1 | Y = 1) = x_1 \cdot P([\mathbf{x}]_1 = 1 | Y = 1) + (1 - x_1) \cdot P([\mathbf{x}]_1 = 0 | Y = 1)$$

where x_1 is the value of $[\mathbf{x}]_1$. The above equation simplifies to our desired values subject to the values x_1 takes. If $x_1 = 1$, then $P([\mathbf{x}]_1 | Y = 1) = P([\mathbf{x}]_1 = 1 | Y = 1)$, and if $x_1 = 0$, then $P([\mathbf{x}]_1 | Y = 1) = P([\mathbf{x}]_1 = 0 | Y = 1)$!

	$P([\mathbf{x}]_1 Y = 1)$	$P([\mathbf{x}]_2 Y = 1)$	$P([\mathbf{x}]_3 Y = 1)$	$\log P([\mathbf{x}] Y = 1) = \sum_{\alpha=1}^3 \log P([\mathbf{x}]_{\alpha} Y = 1)$
Alicia	0.09	0.17	0.30	-5.384
Eve	0.91	0.17	0.30	-3.070
Aiden	0.09	0.17	0.70	-4.537
Fred	0.91	0.83	0.70	-0.637
	$P([\mathbf{x}]_1 Y = -1)$	$P([\mathbf{x}]_2 Y = -1)$	$P([\mathbf{x}]_3 Y = -1)$	$\log P([\mathbf{x}] Y = -1) = \sum_{\alpha=1}^3 \log P([\mathbf{x}]_{\alpha} Y = -1)$
Alicia	0.16	0.24	0.73	-3.574
Eve	0.84	0.24	0.73	-1.916
Aiden	0.16	0.24	0.27	-4.569
Fred	0.84	0.76	0.27	-1.758

We now have all of the pieces. To make the final prediction, we add the log-posterior to each and compute the difference of logs.

	$\log P(Y = 1 \mathbf{x}) \propto \log P(Y = 1) + \sum_{\alpha=1}^3 \log P([\mathbf{x}]_{\alpha} Y = 1)$ (boy)	$\log P(Y = -1 \mathbf{x}) \propto \log P(Y = -1) + \sum_{\alpha=1}^3 \log P([\mathbf{x}]_{\alpha} Y = -1)$ (girl)	Prediction
Alicia	-6.077	-4.267	-1 (girl)
Eve	-3.763	-2.609	-1 (girl)
Aiden	-5.230	-5.262	1 (boy)
Fred	-1.330	-2.451	1 (boy)