Psych 186A

Categorization using nonparametric and parametric methods

One popular application of Bayesian inference is to make inference based on previously observed items. In this tutorial, we'll implement both nonparametric and parametric methods for a categorization task.

Suppose we've observed RT measurements from 50 normal words, and from 50 taboo words in a word recognition experiment. These data can be used to learn the two categories, response to normal words and to taboo words, in the phase of category learning. Now a new observation of RT is given, and the categorization model needs to decide the probability that this measurement is from a normal word.

Input:

RTNorm.mat contains a vector, `NormRT`, which store the RT data for 50 normal words. **RTTaboo.mat** contains a vector, `TabooRT`, which store the RT data for 50 taboo words. Load these MAT files at the beginning of your script.

Nonparametric method

We will use Gaussian kernel method that you implemented on Tuesday to compute two probability distributions: P(RT | Normal words) and P(RT | Taboo words)

Category learning

1. Define the RT space, we make a vector `T = 1:stepsize:1000`, where $\text{stepsize} = 0.1$.
2. For each observed RT in a category, compute the kernel values for all the possible values in the RT space `T`. We use the kernel window size, $h = 20$.
3. Sum up the all kernel values across all observations
4. Normalize $K(T)$, so that it becomes a probability density function (PDF)
5. Repeat step 2-4 for the second category. [Hint: use a for loop to compute the two distributions P(RT | Normal words) and P(RT | Taboo words)].
6. Plot the two category distributions, P(RT | Normal words) and P(RT | Taboo words).

Categorization task

$$P(c \mid x) = \frac{p(x \mid c)P(c)}{\sum_{c} p(x \mid c)P(c)}$$

7. You were informed that five testing trials showed RT = [200, 460, 560, 600, 763], but unfortunately the word list in the testing was lost. Use the Bayes rule above to solve this categorization problem, compute the probability of the corresponding word being a normal word for each trial.

Parametric method

1. We assume that the two categories can be represented using Gaussian distributions center at the prototypical RTs. The parametric method estimates the mean and the variance to form the prototype representation of categories. You can compute sample mean and sample standard deviation using the exemplars. [Hint: use Matlab build-in function mean() and std()].

Estimating a Gaussian

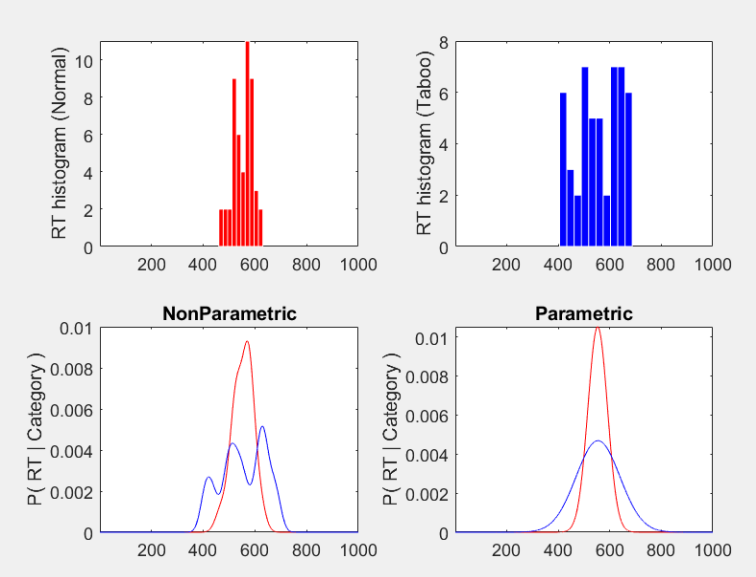
$X = \{x_1, x_2, \ldots, x_n\}$ independently sampled from a Gaussian

$$p(X \mid \mu,\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}$$

maximum likelihood parameter estimates:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

2. Plot the two category distributions, P(RT | Normal words) and P(RT | Taboo words).
3. For the same five testing trials with RT = [200, 460, 560, 600, 763], use the distributions calculated in step 2 to compute the probability of the corresponding word being a normal word for each trial.




```
NonParametric method

prob_cat1val_NonP =

    0.0000    0.3845    0.7517    0.6425    0.0000

Parametric method

prob_cat1_pararst =

    0.0000    0.1678    0.6889    0.5486    0.0000
```