CS 838: Computational Cognitive Science

Final Project: Semi Supervised classification in two dimensions

Priyananda Shenoy (shenoy@cs.wisc.edu)

May 2008

Abstract

We extend the experiment done by Zhu et. al. to determine whether humans do semi-supervised classification to two dimensions. The test subjects classify images into two categories, where the images have two varying features: shape and color. We also try to gain an insight on feature bias: i.e. do users prefer shape more than color in classification.

1 Introduction

Semi supervised learning is the class of machine learning techniques in which both labeled and unlabeled data is used to train the learner, rather than a purely labeled training set used in traditional learning algorithms. An intriguing question which has been studied recently is whether human beings do semi-supervised learning, i.e. do humans make use of both labeled and unlabeled examples in learning a concept. Zhu et. al. gave a positive answer to that question. In [ZHU07], they conducted an experiment on 22 subjects, in which the subjects had to classify objects into two categories, where the objects varied on just one parameter. They were able to measure the shift in the decision boundary caused by showing unlabeled examples to the subjects, and the measured shift matched the mathematical model.

In this project, we aim to extend their experiment to two dimensions, i.e. where the objects have two varying properties. The properties we use are size of the object and the color of the object. We observe the classification boundary before and after showing the subject unlabeled samples, and see if the shift matches our prediction.

2 Ideas and Objectives

At the core of our experiment (and also Zhu et. al.'s experiment) is the notion of *binary classification*. The test subject is shown a visual representation of an object, and asked to assign one of two labels to it. Note that the test subject is not allowed to assign a different label other than the two specified, nor is he/she allowed to make a "fuzzy" assignment to the categories. The subject makes use of the information presented about the object (it's "features") to find a category which best fits the object.

In the experiment conducted by Zhu et. al., the objects varied only in shape. Though not obvious, the shape was controlled by a single parameter, which made the resulting feature space one-dimensional. One thing to note is that even though it was controlled by a single parameter, the visual appearance of the object changed in multiple ways, since the parameter controlled the size, contour and the luminance of the object. Since their experiment worked as expected, the observation we can make is that humans tend to coalesce multiple spatial features into a single feature, even though they might be mathematically independent.

In our experiment, we want to make it obvious that there are two varying features. For this reason, we chose size and color as the two features. This also allows to gain some insight into feature spaces where there is a distinct bias, i.e. humans prefer one feature over another.

The choice of color as one of the features also has the effect of making our experiment more complex. Human perception of color is non-linear (telling colors apart depends on the range) and non-uniform (different people perceive the same object as having different colors). We try to minimize the effects of such perceptual issues as much as possible in our experiment.

We aim to explore these questions in this project:

- What does the decision boundary in a two dimensional feature space look like? Specifically do people prefer color over size in choosing a category? Does the color or the size range matter?
- Does the decision boundary shift after the subject is shown unlabeled data? Is this shift quantifiable, and if yes, can it be predicted?

There are other interesting questions which can be asked in the same setting (which we not address due to lack of time). One question is how the actual colors used matter in our experiment, i.e. does choosing yellow-green instead of red-blue make any difference. There are many other features that can be studied: orientation of the object, lighting effects, other geometric characteristics etc.

3 Experimental methodology

The objects in our experimental setup are filled circles parametrized by two numbers $(size, color)$ where $size \in (0, 100]$ denotes the radius of the circle and $color \in [0, 100]$ denotes the color of the circle (0 = red, 100 = blue).

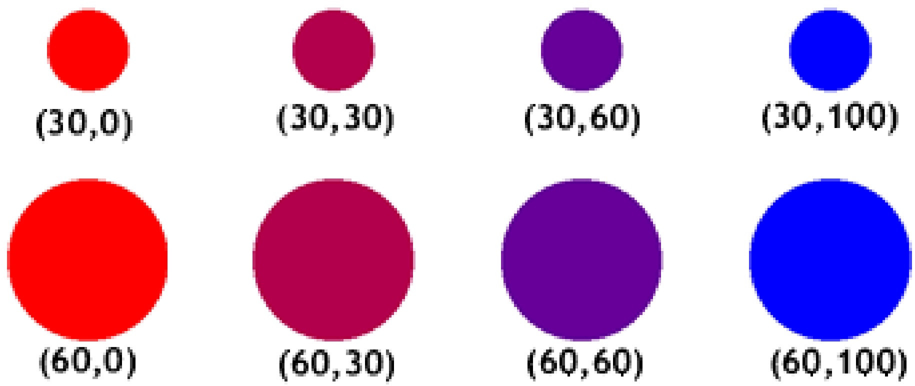

Fig 1: A small portion of our feature space.

Let $\mathcal{A}$ and $\mathcal{B}$ be the two categories. Initially the subject is shown two labeled samples which correspond to initial prototypes: $(30, 0)$ is assigned $\mathcal{A}$ and $(80, 100)$ is assigned $\mathcal{B}$.

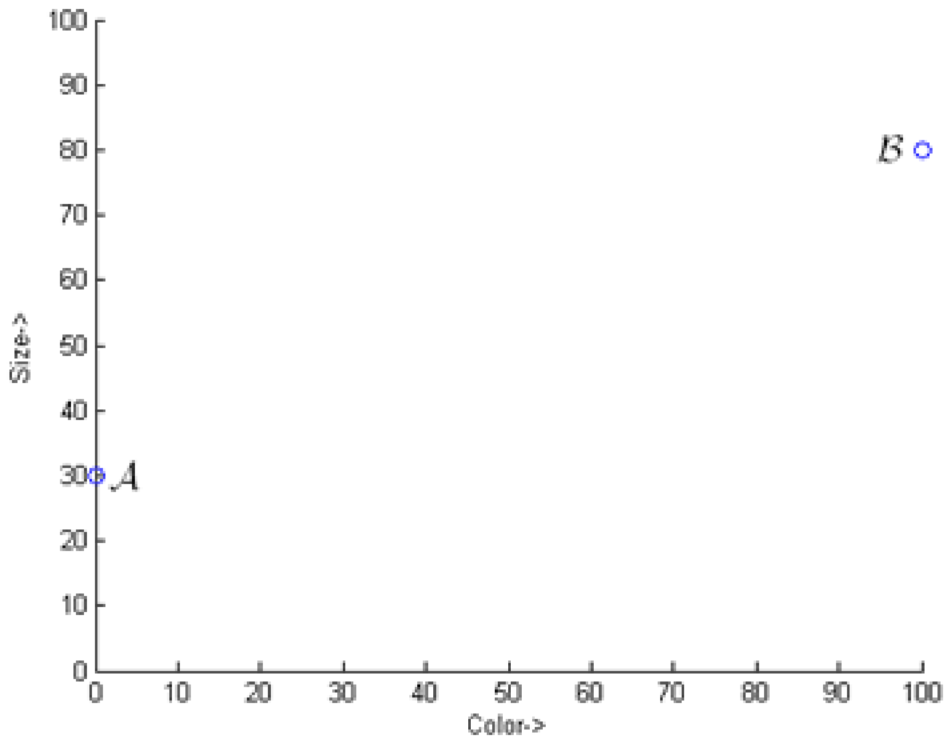

Fig 2: The initial prototypes in our feature space. $(30, 0)$ is assigned label $\mathcal{A}$ and $(80, 100)$ is assigned $\mathcal{B}$.

In Zhu's experiment, there was a natural initial decision boundary which was assumed. In our case, this cannot be assumed. The decision boundary in our two dimensional space is not a point but a curve separating the two regions, a curve of arbitrary complexity and which varies from individual to individual. So instead of assuming the boundary condition, we measure the subject's boundary (approximately). This gives us a chance to examine the subject's feature bias.

The first phase of our experiment presents the subject with 15 unlabeled samples and asks him/her to classify them. There is no time limit, neither is the delay measured. Note that the subject has not been influenced in any way except by the showing of the initial prototypes. These 15 samples are chosen uniformly at random from the entire space. The subject's choices are noted.

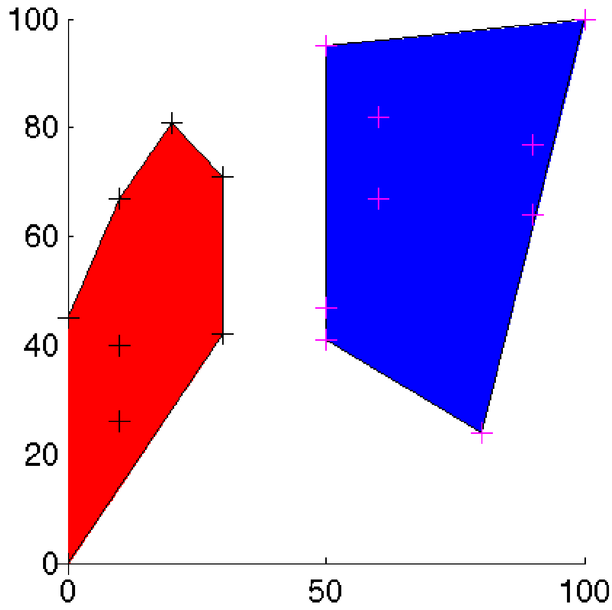

Fig 3: A typical subject's response. The points marked with a black plus were labeled by the subject as $\mathcal{A}$ and those with a pink plus were labeled as $\mathcal{B}$. The red and blue areas are conservative estimates of the areas corresponding to the two labels (captured here by drawing a convex hull around the points).

We now need some way to quantify the feature bias. We will present a simple measure which while approximate gives us a reasonable estimate of the decision boundary. Let L be any line in the feature space. This line divides the feature space into two areas. Arbitrarily assign categories $\mathcal{A}$ and $\mathcal{B}$ to the two areas. Then we define $\Delta(L)$ to be

$$\Delta(L) = \sum_{x \in F} distance(x, L),$$

where F is the set of points

$$F = \{x | \text{label of } x \text{ differs from label of area containing } x \text{ .}\}$$

and $distance(x, L)$ is the perpendicular distance from x to L. Basically we are penalizing points for straying outside their area. Then the *decision boundary* denoted by $\mathcal{L}$ is defined as $\mathcal{L} = arg \min_L \Delta(L)$, i.e. the line which has the least penalty.

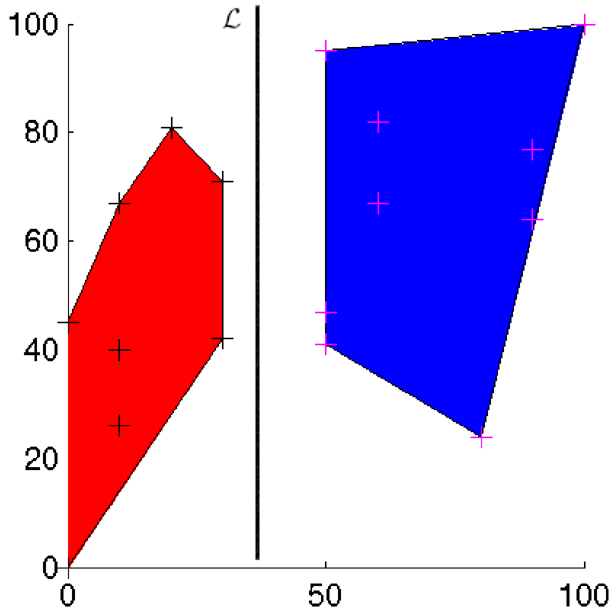

Fig 4: A subject's decision boundary. Note that the decision boundary is not necessarily unique. In this specific case, any 0-penalty line will be a valid decision boundary.

To measure the feature bias of the subject we just take the slope of $\mathcal{L}$. If the slope is 0, i.e. the decision boundary is horizontal, then the subject is completely biased towards size. Similarly if the slope is infinite, i.e the decision boundary is vertical, the subject is completely biased towards color.

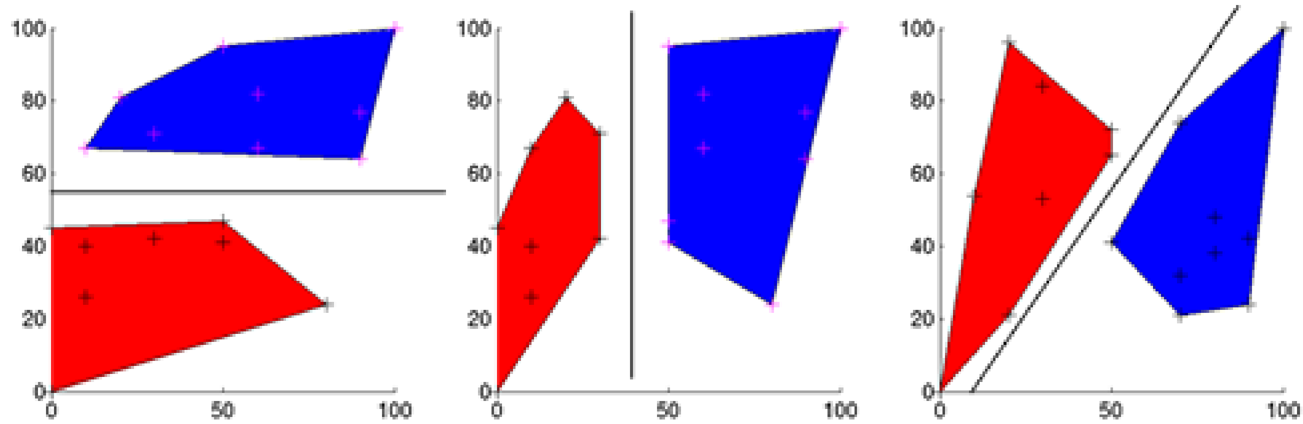

Fig 5: Slopes of different decision boundaries. The left figure shows a decision boundary which is horizontal, i.e. the subject is biased completely towards size. The center figure shows a decision boundary which is vertical, i.e. the subject is biased completely towards color. The right figure shows a decision boundary with some arbitrary slope.

Now that we know the initial decision boundary of the subject, we can proceed with the second phase of the experiment. The subject is shown 20 images each for lasting 5 seconds. The images are unlabeled, and are picked from a normal distribution as explained below.

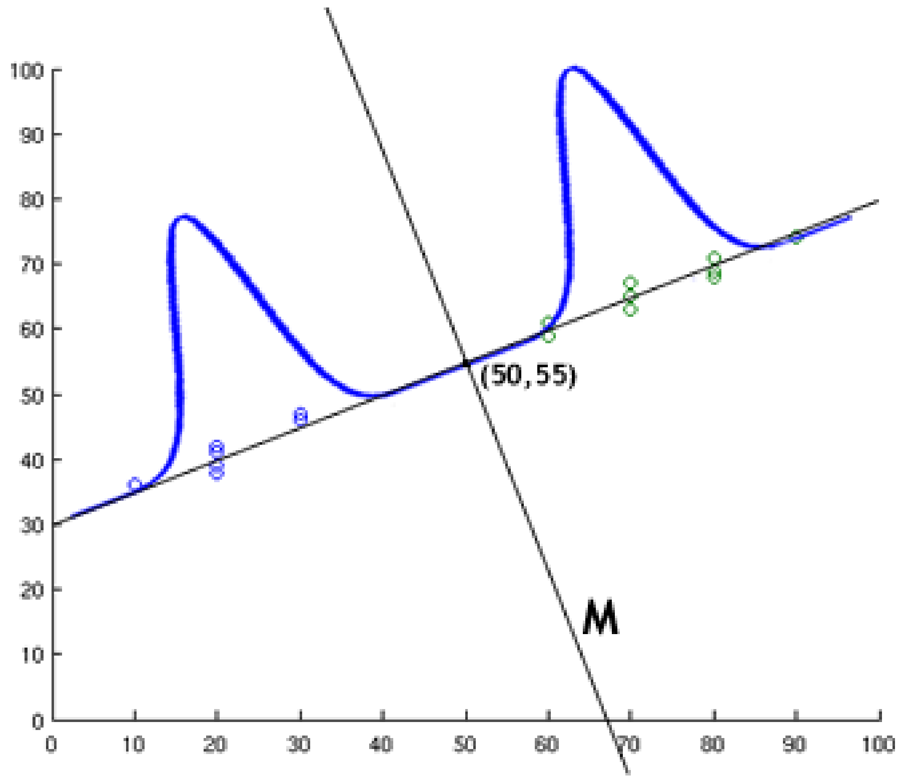

Fig 6: The priming set. We want to shift the subject's decision boundary to the line M.

We want to see if we can shift the subject's decision boundary by showing random unlabeled samples. We pick some arbitrary decision boundary, marked as M in the above picture. We want to shift the decision boundary to M. We generate 10 points each on two one dimensional gaussians, with means -25 and +25 and standard deviation 5, then transform those points onto the line perpendicular to M with it's midpoint $(55, 50)$ being the origin. We use these points as our priming examples.

We then show the subject 15 unlabeled points and ask him/her to classify the points as in Phase 1. We again measure the decision boundary of the subject after priming.

Let us define Δ as $\Delta(M) - \Delta(\mathcal{L})$. Then Δ_0 and Δ_1 be the measurements before and after priming respectively. Then we can define the *shift* as $(\Delta_0 - \Delta_1)$. We predict that the shift is positive, i.e. that the decision boundary is closer to M after priming. [1]

4 Results

We recorded the responses of 14 [2] subjects. Since the data was collected in an uncontrolled environment, we need to qualify the responses to see if they are genuine responses (and not someone giving random responses). For this reason, we call a subject *rational*, if the $\Delta(\mathcal{L}) = 0$. [3] The results for rational users are compiled separately.

[1]This is a rather weak prediction. Since our measure of the decision boundary is a very loose approximation, I can't think of a good mathematical model to predict the shift.

[2]Actually the number of responses was higher than this: 14 is after removing random responses.

[3]The original measure considered for rationality was non-overlapping decision areas. In our case, since we take the decision areas to be convex, this definition is the same as we are using.

Table 1: Recorded responses from test subjects

Test subject responses									
Subject	Before Priming				After Priming				Shift
	S†	$\Delta(\mathcal{L})$	$\Delta(M)$	Δ_0	S	$\Delta(\mathcal{L})$	$\Delta(M)$	Δ_1	
79	-0.80	14.84	15.21	0.37	-0.34	40.71	41.14	0.43	-0.06
77	0.33	2.21	55.45	53.24	0.00	0.00	2.68	2.68	50.56
72	0.00	0.00	33.99	33.99	0.20	0.00	29.52	29.52	4.47
62	∞	0.00	74.24	74.24	1.65	0.00	119.85	119.85	-45.62
54	0.00	0.00	33.99	33.99	-0.30	0.00	5.37	5.37	28.62
66	∞	0.00	86.76	86.76	∞	0.00	44.72	44.72	42.04
58	-0.61	3.90	12.52	8.63	-0.57	0.00	26.83	26.83	-18.21
73	0.00	0.00	33.99	33.99	0.04	4.20	18.78	14.59	19.40
81	0.00	0.00	32.20	32.20	0.00	0.00	28.62	28.62	3.58
61	0.45	0.00	39.35	39.35	0.00	0.00	2.68	2.68	36.67
78	0.03	10.00	32.20	22.20	0.04	4.20	18.78	14.59	7.62
64	0.10	2.99	67.08	64.10	0.00	0.00	28.62	28.62	35.48
65	∞	0.00	44.72	44.72	∞	0.00	86.76	86.76	-42.04
56	1.67	0.00	119.85	119.85	∞	0.00	44.72	44.72	75.13
75	0.00	0.00	25.04	25.04	0.00	0.00	2.68	2.68	22.36
80	0.03	1.50	67.08	65.58	0.00	0.00	28.62	28.62	36.96
52	-0.36	0.00	13.42	13.42	0.00	0.00	28.62	28.62	-15.21

† Slope of $\mathcal{L}$.

5 Conclusion

In $\approx 70\%$ of the cases, the result was as expected: the shift was positive. However for 30% of the subjects the measured shift didn't match the prediction. Several reasons might have caused this: the experimental setup was not correct, the duration of the priming was not sufficient etc. The sample size is quite small, and drawing inferences from it might lead to an incorrect hypothesis.

References

[ZHU07] X. Zhu, T. Rogers, R. Qian, C. Kalish. Humans Perform Semi-Supervised Classification Too. In *Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, 2007.