I acknowledge that this homework is solely my effort. I have done this work by myself. I have not consulted with others about in any way. I have not received outside aid (outside of my own brain). I understand that violation of these rules contradicts class policy on academic integrity.

Rafael Antonio Garcia Mar

Rafael Garcia

Jan 31, 2020

Homework 1

Part A: Theory

Maximum Posterior vs Probability of Chance

We know the probabilities add to 1:

$$\sum_{i=1}^{c} P(w_i|\vec{x}) = 1$$

Let's assume that $P(w_{max}|\vec{x}) < \frac{1}{c}$, then $P(w_i|\vec{x}) < \frac{1}{c}$. Then

$$\sum_{i \neq max} P(w_i|\vec{x}) = 1 - P(w_{max}|\vec{x})$$
$$> 1 - \frac{1}{c}$$

Assuming $P(w_i|\vec{x}) = P(w_j|\vec{x})$, $i,j \neq max$ $\Rightarrow$ $P(w_i|\vec{x}) > \frac{c-1}{c}\frac{1}{c-1} = \frac{1}{c} > P(w_{max}|\vec{x})$

Which is a contradiction.

Then $P(w_{max}|\vec{x}) \geq \frac{1}{c}$.

The probabity of making an error is

$$P(error) = \int \sum_{i \neq max} P(w_i|\vec{x})P(\vec{x})\,d\vec{x} = \int [1 - P(w_{max}|\vec{x})]P(\vec{x})\,d\vec{x}$$

If we use Bayes rule to make a decision then $P(w_{max}|\vec{x}) \geq \frac{1}{c}$

$$P(error) = \int [1 - P(w_{max}|\vec{x})]\,P(\vec{x})\,d\vec{x}$$
$$\leq \left(1 - \frac{1}{c}\right)\int P(\vec{x})\,d\vec{x}$$
$$= \frac{c-1}{c}$$

Bayes decision rule classifier

Choose w_1 if $P(w_1)P(\vec{x}|w_1) > P(w_2)P(\vec{x}|w_2)$.

Using the priors $P(w_1) = P(w_2) = \frac{1}{2}$:

Choose w_1 if $P(\vec{x}|w_1) > P(\vec{x}|w_2)$.

Using the independence of $\vec{x}$.

Choose w_1 if $\prod_{i=1}^{d} P(x_i|w_1) > \prod_{i=1}^{d} P(x_i|w_2)$

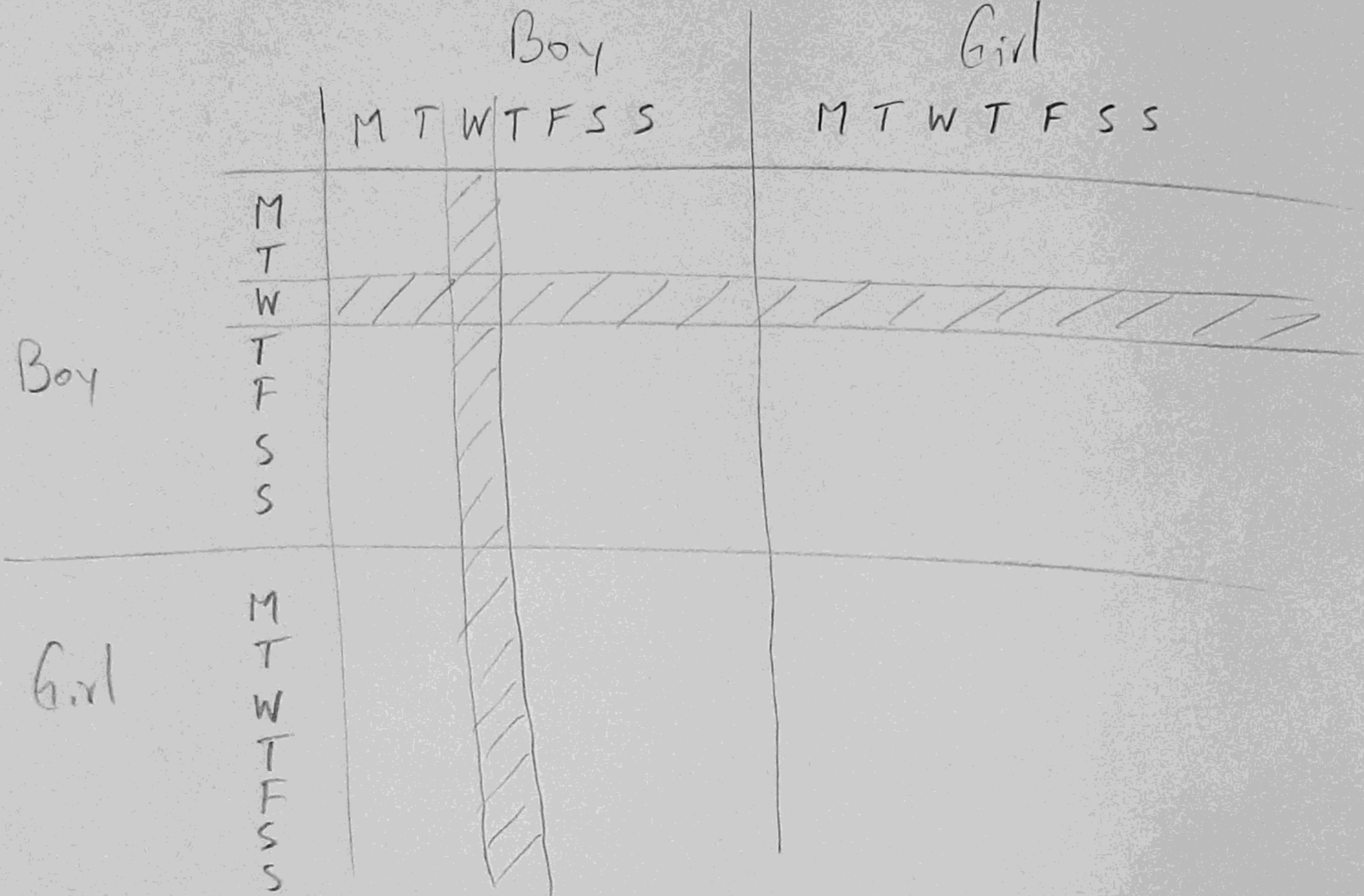
Choose w_1 if $\prod_{i=1}^{d} p^{x_i}(1-p)^{1-x_i} > \prod_{i=1}^{d} (1-p)^{x_i} p^{(1-x_i)}$

Choose w_1 if $\sum_{i=1}^{d} x_i \log p + (1-x_i)\log(1-p) > \sum_{i=1}^{d} x_i \log(1-p) + (1-x_i)\log p$

Choose w_1 if $\sum_{i=1}^{d} x_i \cancel{\log\frac{p}{1-p}} > \sum_{i=1}^{d} (1-x_i)\cancel{\log\frac{p}{1-p}}$

Choose w_1 if $\sum_{i=1}^{d} x_i > \frac{d}{2}$.

The Ditzler household growing up



The size of the probability space with at least 1 boy is 27. The size of the probability space with 2 boys given at least 1 boy is 13. Then

$$P(2 \text{ boys} \mid \text{at least } 1 \text{ boy}) = \frac{13}{27}$$

Linear classifier with a margin

$\vec{x}_1 \in C_1 \quad (y_1 = 1)$

$\vec{x}_2 \in C_2 \quad (y_2 = -1)$

Constraints: $\vec{w}^T \vec{x}_1 + b = 1$

$\vec{w}^T \vec{x}_2 + b = -1$

Minimize: $\|\vec{w}\|_2^2$

Lagrange multipliers:

$$\mathcal{L} = \|\vec{w}\|_2^2 - \lambda_1 f_1(\vec{w}) - \lambda_2 f_2(\vec{w}) \leftarrow \text{Lagrangian}$$

where

$f_1(\vec{w}) \equiv \vec{w}^T \vec{x}_1 + b - 1 = 0$

$f_2(\vec{w}) \equiv \vec{w}^T \vec{x}_2 + b + 1 = 0$

Euler-Lagrange equations:

$$\frac{\partial \mathcal{L}}{\partial \vec{w}^T} = \vec{w} - \lambda_1 \vec{x}_1 - \lambda_2 \vec{x}_2 = 0 \quad (1), \qquad \frac{\partial \mathcal{L}}{\partial b} = -\lambda_1 - \lambda_2 = 0 \quad (2)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = -\vec{w}^T \vec{x}_1 - b + 1 = 0 \quad (3) \checkmark$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = -\vec{w}^T \vec{x}_2 - b - 1 = 0 \quad (4) \checkmark$$

constraints are satisfied

$$(1) \Rightarrow \vec{w} = \lambda_1 \vec{x}_1 + \lambda_2 \vec{x}_2, \quad (2) \Rightarrow \lambda_1 = -\lambda_2 \quad \Rightarrow \quad \vec{w} = \lambda(\vec{x}_1 - \vec{x}_2) \quad (5)$$

(3) and (4) $\Rightarrow$

$$\vec{w}^T(\vec{x}_1 + \vec{x}_2) = -2b$$
$$\vec{w}^T(\vec{x}_1 - \vec{x}_2) = 2 \quad (6)$$

(5) and (6) => $\vec{W}^T\vec{W} = 2\lambda$, $\lambda = \frac{1}{2}|\vec{W}|^2$ (7)

(7) and (5) => $\frac{\vec{W}}{|\vec{W}|^2} = \frac{\vec{X}_1 - \vec{X}_2}{2}$

$$|\vec{W}|^{-1} = \frac{1}{2}|\vec{X}_1 - \vec{X}_2|$$

$$\Rightarrow \quad \vec{W} = 2\,\frac{\vec{X}_1 - \vec{X}_2}{|\vec{X}_1 - \vec{X}_2|^2}$$

b can be obtained from (3) or (4):

$$b = -\vec{W}^T\vec{X}_1 + 1$$

$$b = -2\,\frac{\vec{X}_1^T - \vec{X}_2^T}{|\vec{X}_1 - \vec{X}_2|^2}\,\vec{X}_1 + 1$$

Then, two points are enough to determine the location of the maximum margin hyperplane.

Decision making with Bayes

* Generative models: $\frac{P(\vec{x}|w)P(w)}{P(\vec{x})}$

  - Use data to model the distribution of data in order to categorize the data.
  - Generative models can generate new data instances.
  - Generative models try to learn the process of how data is generated.

* Discriminant models: $P(w|\vec{x})$

  - Don't care about the process that generated the data.
  - Generally give better performance in classification tasks.

Knowing $P(\vec{x})$ for generative models might be useful since it is the marginal probability that an observation $\vec{x}$ is seen.