Mathematical Notation

This sheet has some useful notation which you will see throughout the course.

Individuals / Examples

- We typically say we have n individuals and denote a representative individual as i such that we say there are individuals $1, ..., i, ..., n$ or $i = 1, ..., n$.

Variables

- An input variable/features: x

- An input variable/feature when you have more than one. Note that we've used k here to note the number of input variables but this is quite flexible and people often use m, p and d as well: $x_1, x_2, ..., x_k$.

- Sometimes, for brevity, you can write the input features as a vector x or X were this contains all of $x_1, x_2, ..., x_k$. This just simplifies notation. In this case we'd say $x \in R^k$ i.e. x is k-dimensional.

- Target variable/class: y

- In machine learning you generally use superscript notation to denote the specific value for an individual i. So you would have input variables $x_1^{(i)}, x_2^{(i)}, ..., x_k^{(i)}$ and target variable $y^{(i)}$. You may sometimes see the i as a subscript as well.

Datasets

- A dataset has n individuals, k input features and possible a target variable. We write datasets with the following notation: $\{(x^{(1)}, y^{(1)}), ..., (x^{(n)}, y^{(n)})\}$ where x is a vector of all input features. If you didn't have a target features then the dataset would be $\{(x^{(1)}), ..., (x^{(n)})\}$.