Supervised learning $\iff$ **labeled data**. If we have some data point (x,y) x is the instance (or example) and y is the label. $x \in$ Instance Space $y \in$ Label Space. The **Concept Space** is the set of functions that the true function lives in. $H =$ **Hypothesis Space** is the set of functions that we search over.

(someone please validate the above paragraph)

Entropy

$H = -\sum_{i=1}^{n} P(p_i) * \log(p_i)$

Online Learning

We get data points one at a time and cannot revisit them. Gives: **mistake bound algorithms** which are algorithms that will make a finite number of mistakes while in use. Note that to be useful we often want the mistake bound to be better than $O(n)$ on the number of datapoints (otherwwise they're obviously not very useful).

Linear Classifiers

Find a hyperplane to split the data. This is notationally easy as we can define the classifier as:

$y' = \text{sign}(w^T.x + b)$ (can roll the bias in as a feature, then it's: $y' = \text{sign}(w^T.x)$)

Perceptron Alg

if $y \neq y'$ then $w = w + r(y_i x_i)$

mistake bound: $t \leq \frac{R^2}{\gamma^2}$ where t is the number of mistakes the algorithm will make, R is the radius of the dataset (furthest distance from the origin to a point in the set), and γ is the distance between the point nearest the optimal hyperplace.