Lecture 1 Mathematical Foundations of Machine Learning

Elements of Machine Learning

Imagine we want to take a photograph of a face and determine whether the person is smiling or not.

Key idea: we represent whether face is smiling with a <u>model</u> - a mathematical description of the data

Basis steps:

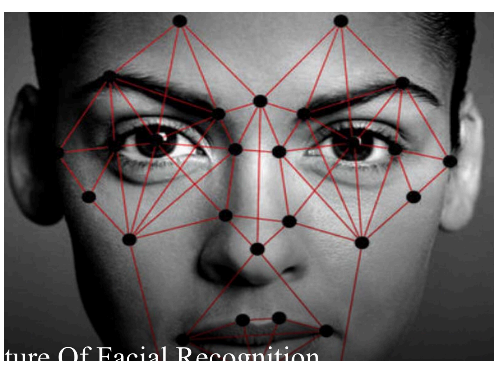
① Collect raw data — e.g. photographs of faces

② preprocessing – change the data to simplify subsequent operations without losing relevant information — e.g. crop images to only contain <u>one</u> face

Center face

resize so all images have same # pixels

③ Feature extraction – reduce raw data by extracting features or properties relevant to model (This step if often unnecessary in modern image recognition systems based on deep neural nets)



- e.g. distances between pairs of facial landmarks

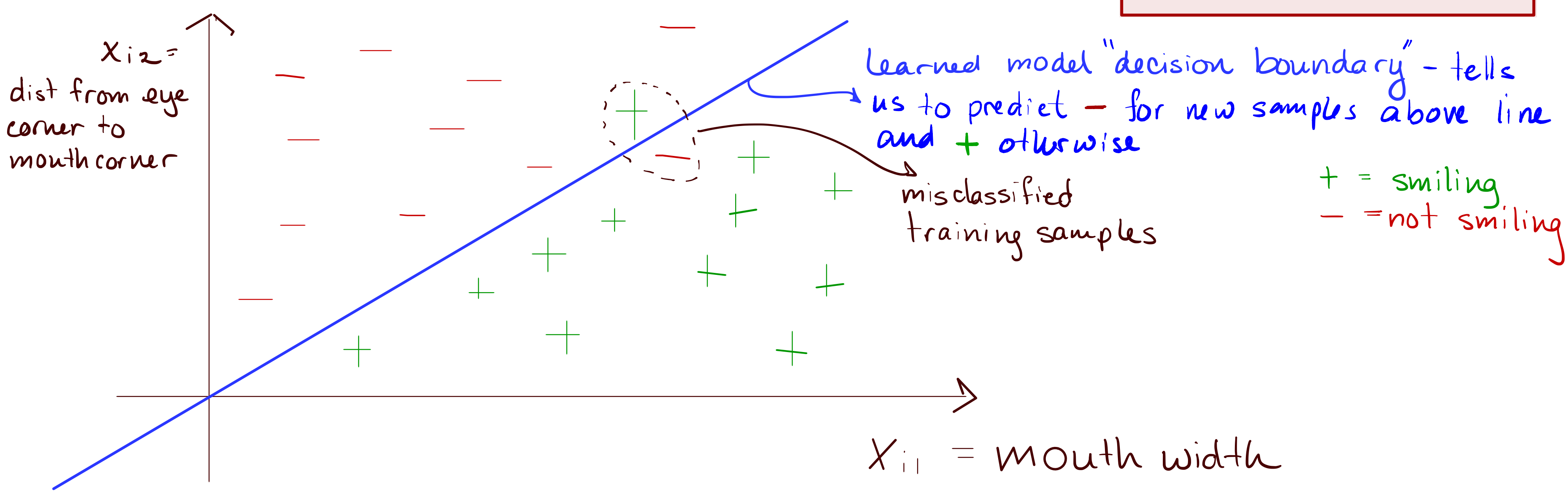
④ generate training samples = large collection of examples we can use to learn a model

$(\underline{x}_i, y_i)$ for $i = 1, \ldots, n$ ⇒ n = # training samples
y_i = i^{th} sample's label
$\underline{x}_i$ = i^{th} sample's features
x_{ij} = j^{th} feature of i^{th} sample

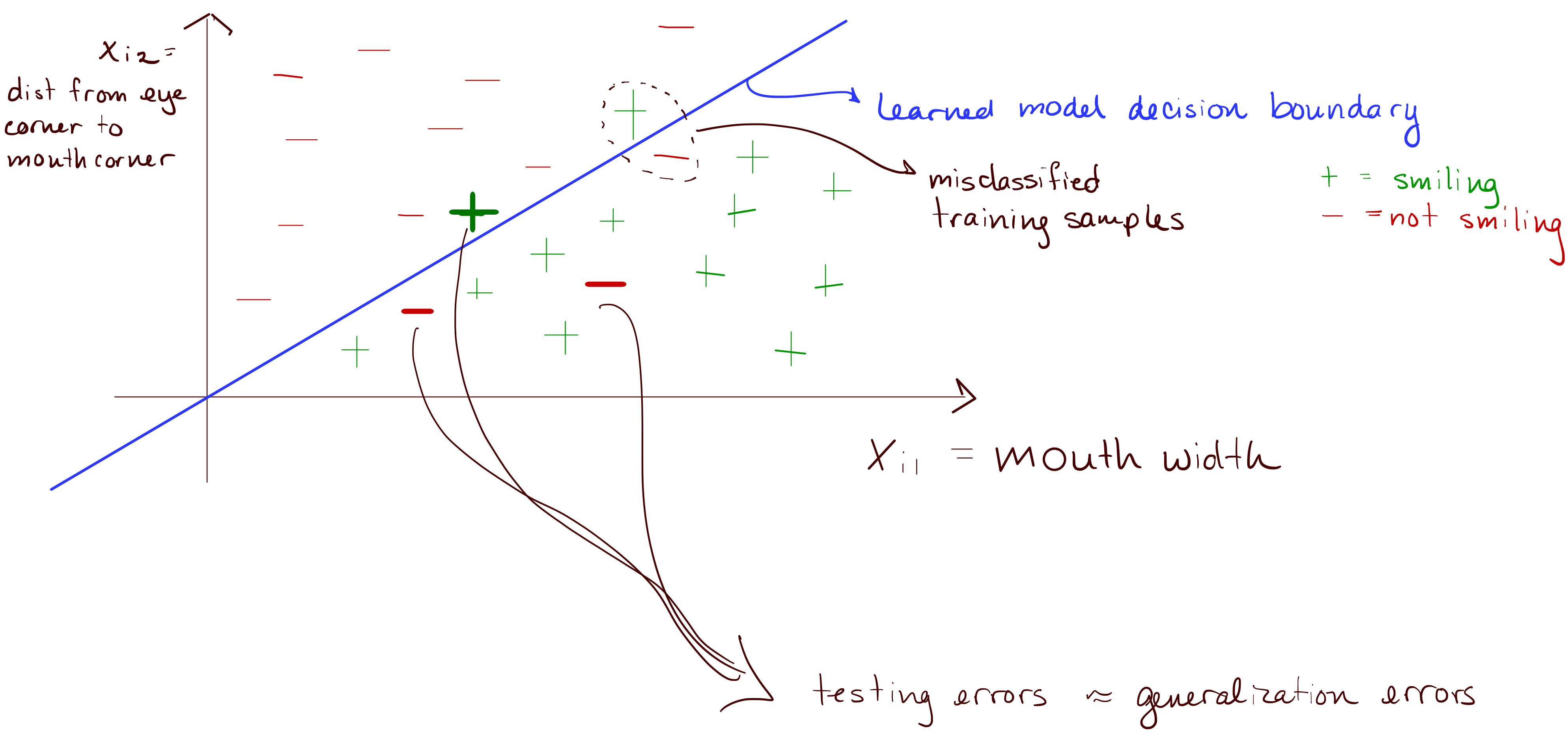
⑤ To learn model, we choose a loss function = a measure of how well a model fits data – e.g. % of samples misclassified as "smiling"

⑥ Finally, learn the model – search over collection of candidate models or model parameters to find one that minimizes loss on training data

some models learn new "features" from original input features



⑦ Characterize generalization error — error of our prediction on new data that was not used for training. Sometimes we estimate this using a set of test samples not used for training



Our First ML Problem — Classification

Learning we observe training data $(\underline{x}_i, y_i)$ for $i = 1, , n$

where $\underline{x}_i \in \mathbb{R}^p$ is a vector of p real numbers called the feature

and $y_i \in \mathbb{R}$ or $y_i \in \{-1, +1\}$ or $y_i \in \{0, 1\}$ is the label

Our goal learn a model that predicts a label $\hat{y}$ given a feature vector $\underline{x}_0$

Ex linear model $\hat{y} = w_1 x_{01} + w_2 x_{02} + \quad w_p x_{0p}$

$w_1, \ , w_p$ = weights to be learned from data

> This is tedious to write!
> Let's use some shorthand
> that will make many things easier

Let $\underline{w} = \begin{bmatrix} w_1 \\ w_2 \\ \\ w_p \end{bmatrix} \in \mathbb{R}^p$ be the weight vector

real numbers

Let $\underline{x}_0 = \begin{bmatrix} x_{01} \\ x_{01} \\ \vdots \\ x_{0p} \end{bmatrix} \in \mathbb{R}^p$ be the feature vector

Then our linear model can be equivalently written as

$$\hat{y} = \underbrace{\langle \underline{w}, \underline{x}_0 \rangle}_{\text{Inner product of two vectors}} = \underline{w}^T \underline{x}_0 = [w_1, \quad , w_p] \begin{bmatrix} x_{01} \\ x_{01} \\ \vdots \\ x_{0p} \end{bmatrix} = \underline{x}_0^T \underline{w} = \langle x_0, w \rangle$$

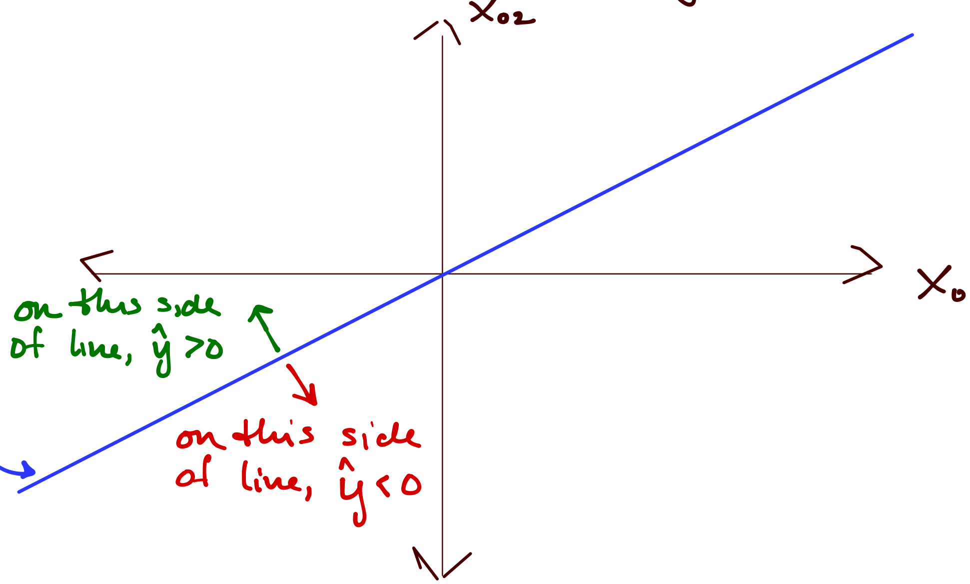
vector transpose

$\boxed{\text{Ex}}$ $p = 2$ $\quad \underline{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ When is $\hat{y} = \langle \underline{w}, \underline{x}_0 \rangle > 0$ and when is $\hat{y} < 0$?

$$\langle \underline{w}, \underline{x}_0 \rangle = w_1 x_{01} + w_2 x_{02} > 0$$

$$\Rightarrow -2 x_{01} + x_{02} > 0$$

$$\Rightarrow x_{01} < x_{02}/2$$



Line = set of $\underline{x}_0$ where $\hat{y} = 0$