ISTA 421 + INFO 521
Introduction to Machine Learning

Lecture 2: Linear Models

Clay Morrison
claytonm@email.arizona.edu
Harvill 437A
Phone 621-6609

22 August 2018

SISTA 1

Today

- Introduction
- A simple learning problem
- Linear Model (what is the *model*)
- Loss Function (what is a *good* model)
- Least Squares (finding the '*best*' model)
- Prediction
- Moving to higher dimensions

SISTA 2

What is Machine Learning?

- The goal of machine learning is to build computer systems that can adapt and learn from their experience. (Dietterich, 1999)
- Machine learning usually refers to changes in systems that perform tasks associated with artificial intelligence. Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. (Nilsson, 1996)
- Some reasons for adaptation:
  – Some tasks can be hard to define except via examples
  – Adaptation can improve a human-built system, or track changes over time
- Goals can be autonomous machine performance, or enabling humans to learn from and understand data (data mining and modeling)

Ack: this and some following content adapted from Chris Williams 2006

Some of the Roots of Machine Learning

- **Philosophy**: epistemology, philosophy of science, logical inference: *the Problem of Induction*
- **Mathematics**: structure, operations, optimization
- **Physics**: statistical mechanics
- **Statistics**: statistical inference, frequentist & Bayesian
- **Psychological** models (of learning and development)
- **Brain** models, e.g. neural networks
- **Artificial Intelligence**: e.g., discovering rules using decision trees, inductive logic programming, autonomy
- **Engineering**: Statistical pattern recognition, operations research, adaptive control theory

Typical Supervised Machine Learning Workflow

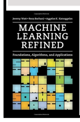
(0) **Define the problem.** What is the task we want to teach a computer to do?

(1) **Collect data.** Gather data for training and testing sets. The larger and more diverse the data the better.

(2) **Design features.** What kind of features best describes the data?

(3) **Train the model.** Tune the parameters of an appropriate model on the training data using numerical optimization.

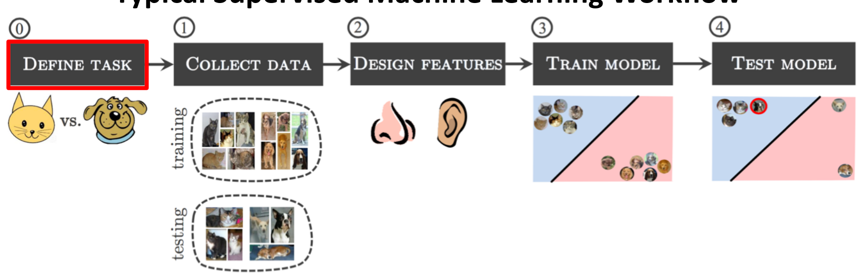
(4) **Test the model.** Evaluate the performance of the trained model on the testing data. If the results of this evaluation are poor, re-think the particular features used and gather more data if possible.

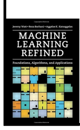


From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

5

Typical Supervised Machine Learning Workflow

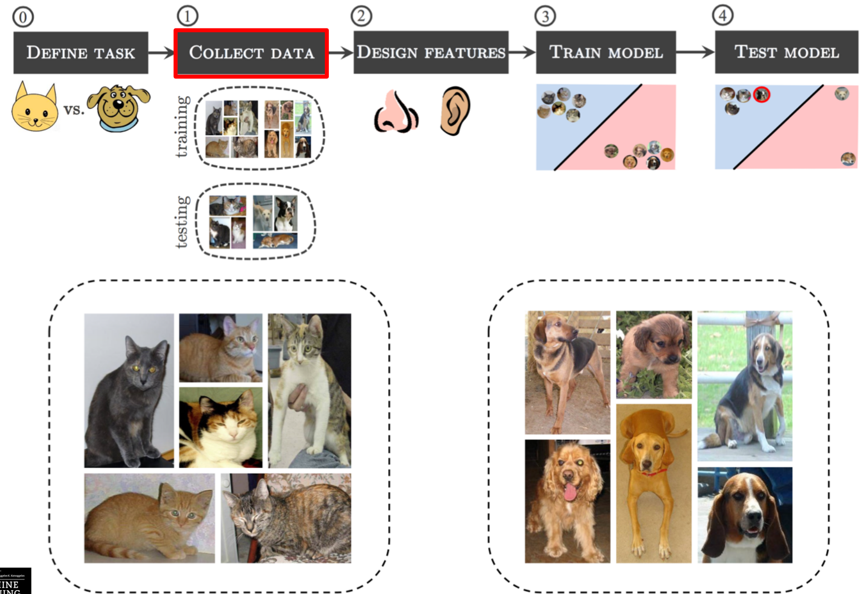




From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

6

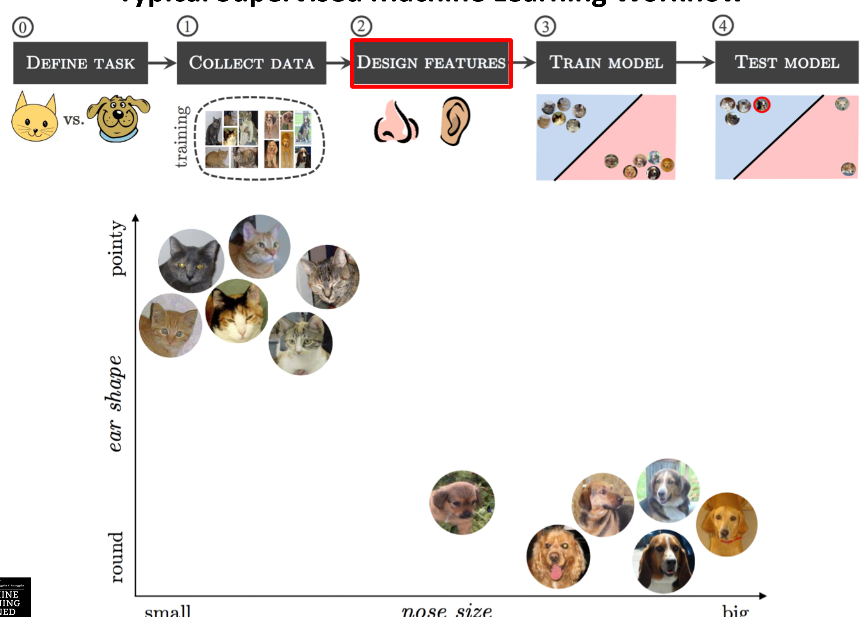
Typical Supervised Machine Learning Workflow



From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

7

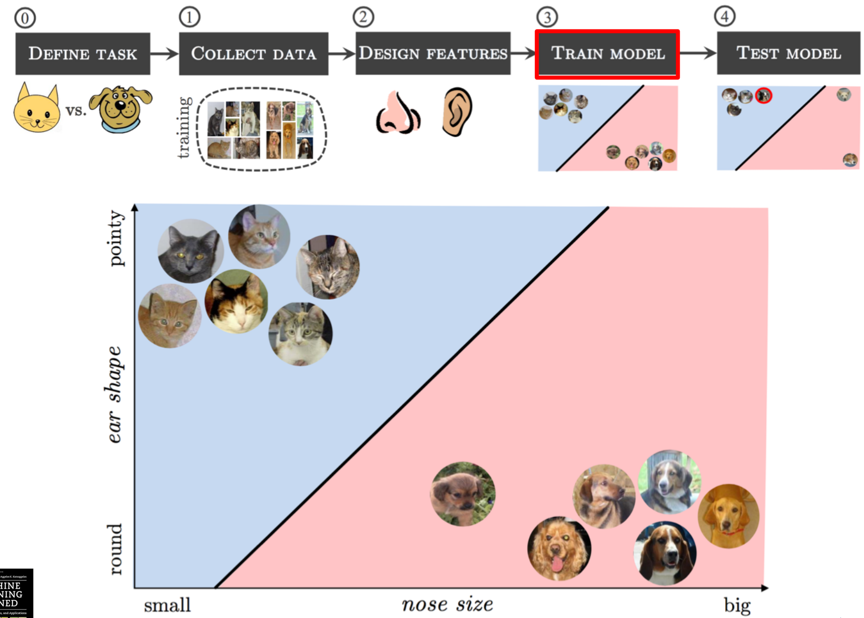
Typical Supervised Machine Learning Workflow

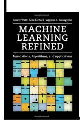


From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

8

Typical Supervised Machine Learning Workflow

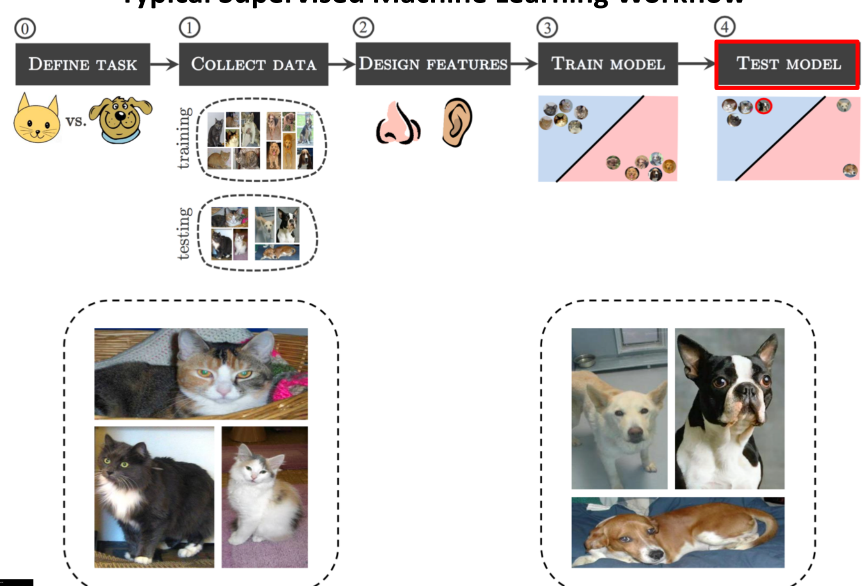


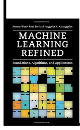


From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

9

Typical Supervised Machine Learning Workflow

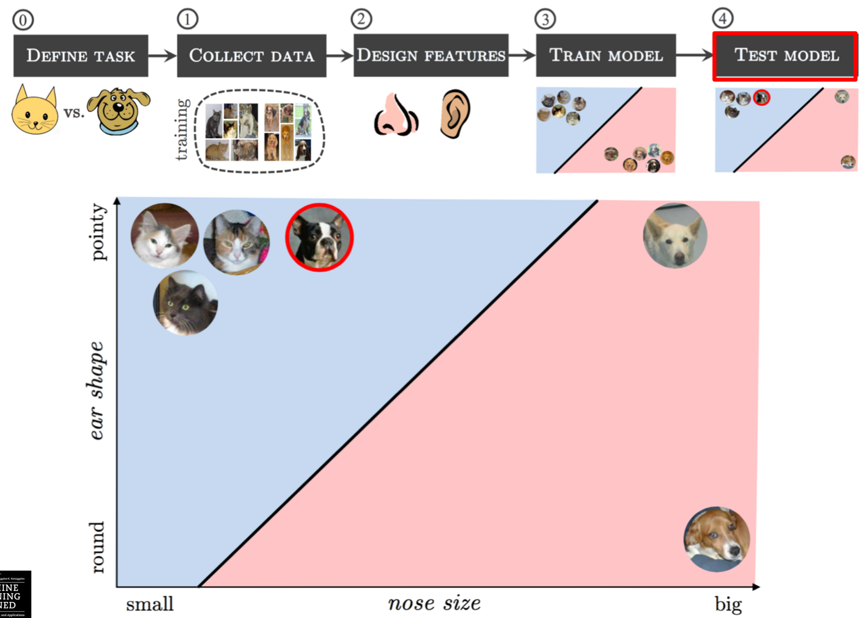


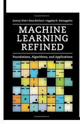


From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

10

Typical Supervised Machine Learning Workflow

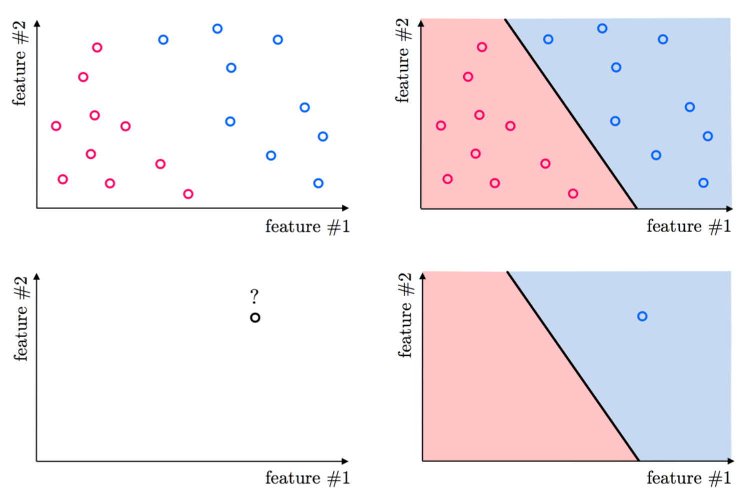




From Watt, Borhani & Katsaggelos (2016) *Machine Learning Refined*
http://mlrefined.wix.com/home-page

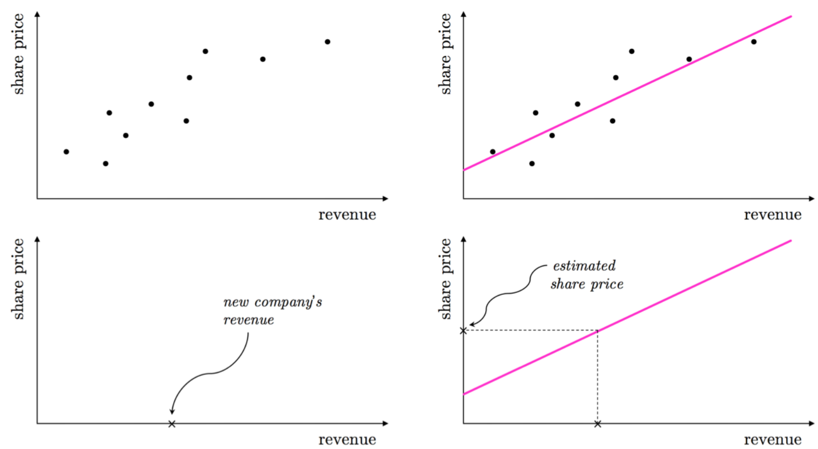
11

Classification



12

Regression



Three General Classes of ML

- **Supervised learning** – model $p(y|\mathbf{x})$
  - Given model and data with correct output (label)
  - Regression, Classification, etc.
- **Unsupervised Learning** – model $p(\mathbf{x})$
  - Only given input data (no output)
  - Clustering, Latent Models, Projection methods, etc.
- **Reinforcement Learning** – model $p(\mathbf{s}_{t+1}|\mathbf{s}_t, a)$
  - Given input data choose output (action), get *grade* for output
  - Learning to choose better actions
  - Markov decision processes, POMDPs, planning

Supervised Learning: Some Terminology

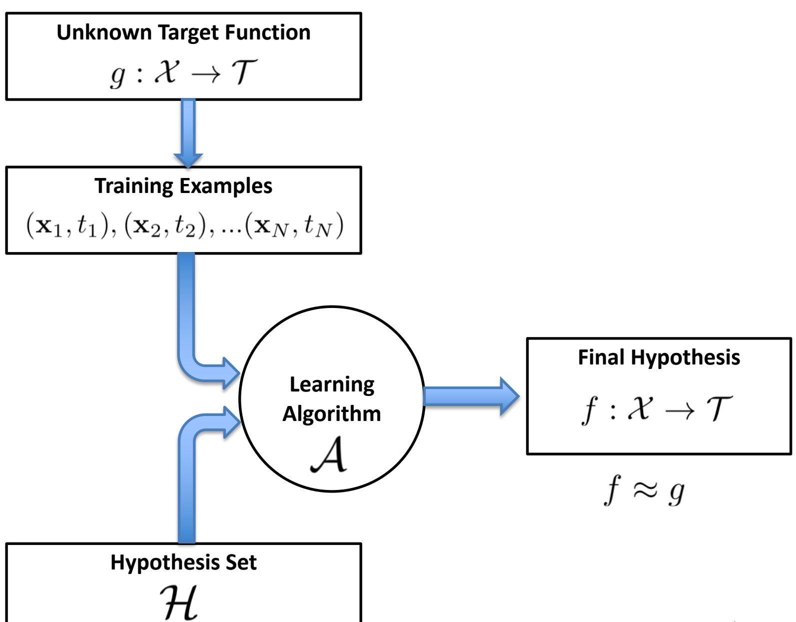
- Input: $\mathbf{x}$ (customer application)
- Output: t (good/bad customer?)
- Target function: $g : \mathcal{X} \rightarrow \mathcal{T}$ (*ideal* credit approval fn)
- Data: $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), ...(\mathbf{x}_N, t_N)$ (historical records)

$\Big\downarrow$ *learning*

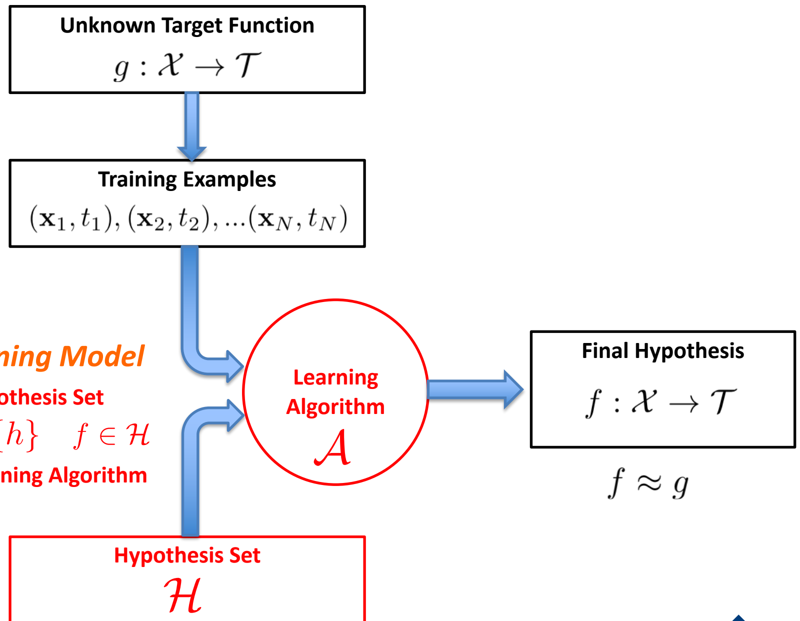
- Hypothesis: $f : \mathcal{X} \rightarrow \mathcal{T}$ (formula to be used)

Adapted from Yaser S. Abu-Mostafa et al., *Learning from Data*

15



16



The ***Learning Model***

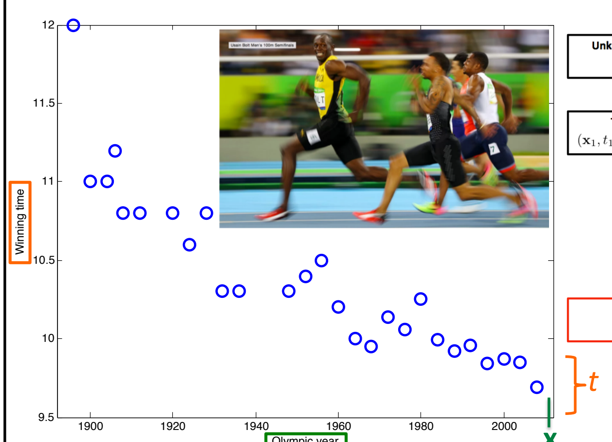
1) The **Hypothesis Set**

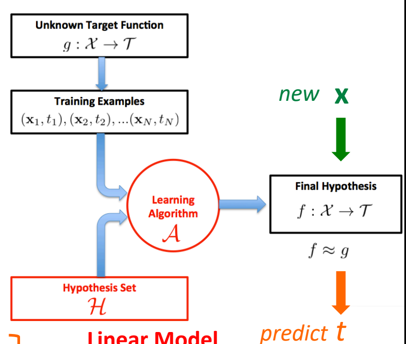
$\mathcal{H} = \{h\} \quad f \in \mathcal{H}$

2) The **Learning Algorithm**

17

A simple learning problem





Want to describe **Winning time** (*t*) as a function of **Olympic year** (**x**)

18

Defining a Model

- Define function that maps inputs (Olympics year, x_i) to output or target values (Winning times, t_i)

  $t = f(x)$

- The model itself likely has parameters, which we'll generically refer to as 'θ' here. It is common to make them explicit within a function:

  $t = f(x\,;\,\theta)$

19

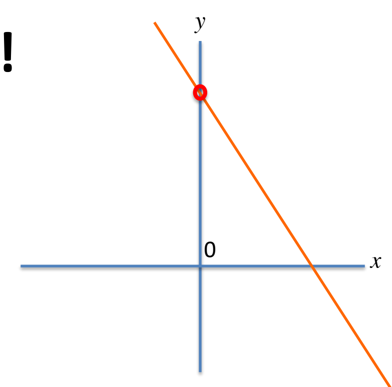
Lines!

- **Slope-intercept form**

  $$y = \underline{m}x + \underline{b}^*$$

- General (standard) form

  $$ax + by + c = 0$$



slope $m = -\frac{a}{b}$

y-intercept $b^* = -\frac{c}{b}$

x-intercept $= -\frac{c}{a}$

20

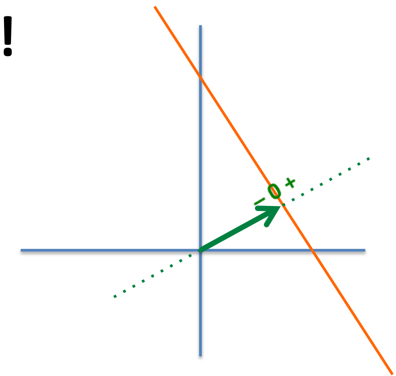
Lines!

- **General (standard) form**

$$\underline{a}x + \underline{b}y + \underline{c} = 0$$

- Slope-intercept form

$$y = \underline{m}x + \underline{b^*}$$



slope $m = -\frac{a}{b}$

y-intercept $b^* = -\frac{c}{b}$

x-intercept $= -\frac{c}{a}$

21

Linear Relationship

- ***y* = *mx* + *b*** **(or *t* = w_1x + w_0)**
  - the classic line (in 2D space)
  - For a given line, *m* and *b* are the ***parameters*** and *x* is a ***variable*** in the relationship: $y = f(x\ ;\ m, b)$
  - When considering alternate lines, we are adjusting *m* and *b*
- Generally, as long as the values that vary (assuming the others are constant) are not themselves involved in anything more than
  (1) addition and
  (2) scalar multiplication,
  … then the relationship is ***linear***.

Let's consider the relationship between *y* and *x*

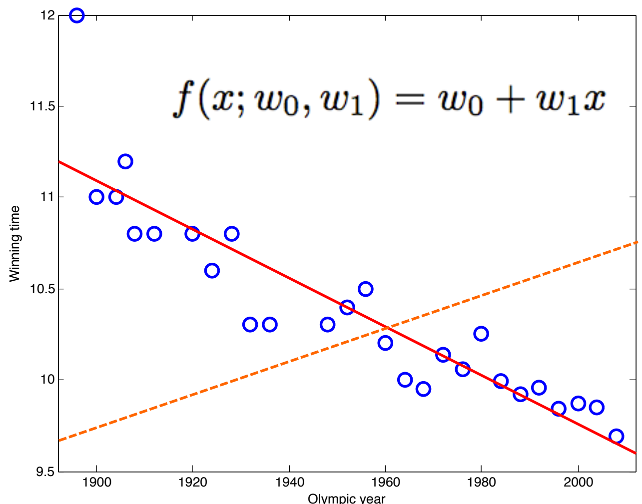
$y = mx^2 + c$ $\quad y = \sin(x)$ $\quad \sqrt{y} = mx + c$ Not linear rel. btwn *x,y*

$y = mx + c^2$ $\quad y = x\sin(m) + c$ Is linear rel. btwn *x,y*

What about the relationship between *y* and *m* (*m* is a parameter!)

22

Data with line (particular w_0 & w_1)



(The red line happens to be a "best" fit)

(The dashed orange line does not describe the trend in the data very well; not a good fit)