Introduction to Data Science

**What is Machine Learning?**

* Learning is any process by which a system improves performance from experience
* Algorithms that can figure out how to perform a task from exmaples
* Learning **task T** is any process by which a system improves performance **measured by P** from experience **E.**
  + Example:
    - **T:** SPAM-filtering
    - **P:** % of SPAM/HAM-mails filtered out
    - **E:** emails that are labelled as SPAM/HAM

**Advantages of Machine Learning?**

* Can automatically adapt to changes and individual users
* Discover new knowledge
* Mimic human intelligence
* Human engineering is sometimes too difficult

**Types of Machine Learning**

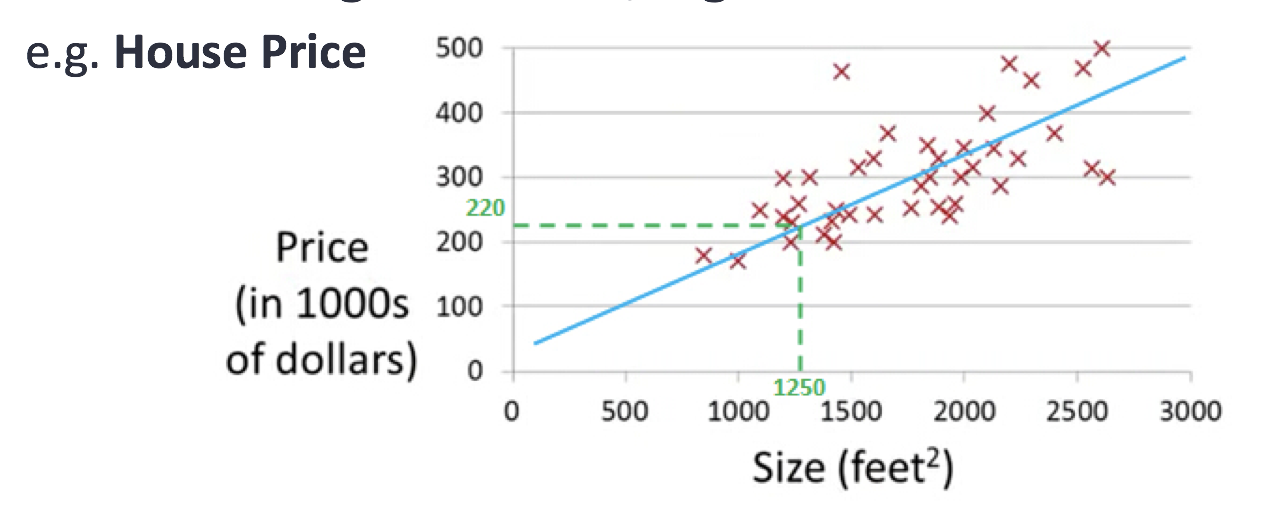
****

**Linear Regression**

Regression:

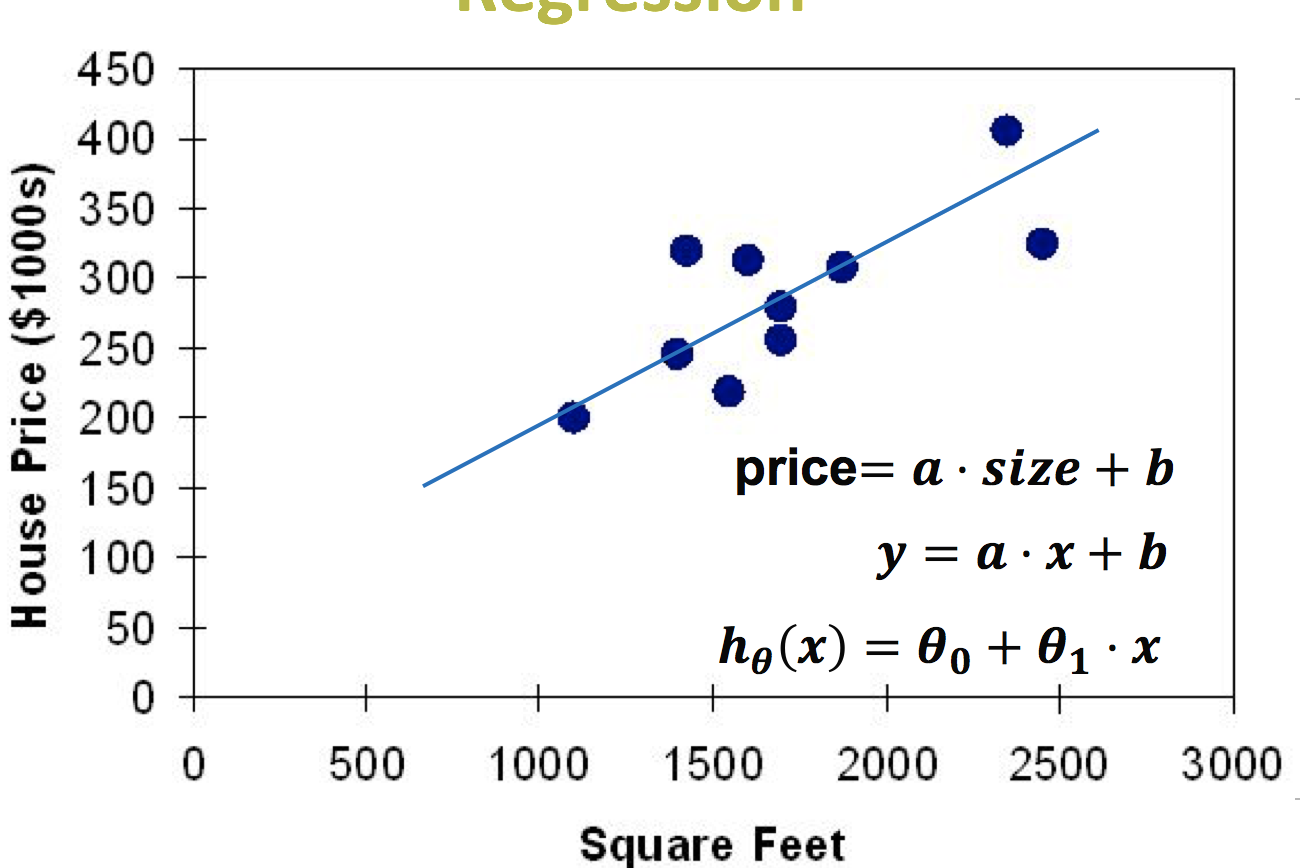
* Estimate the relationship between variables, **using information about examples**

Estimate a target variable

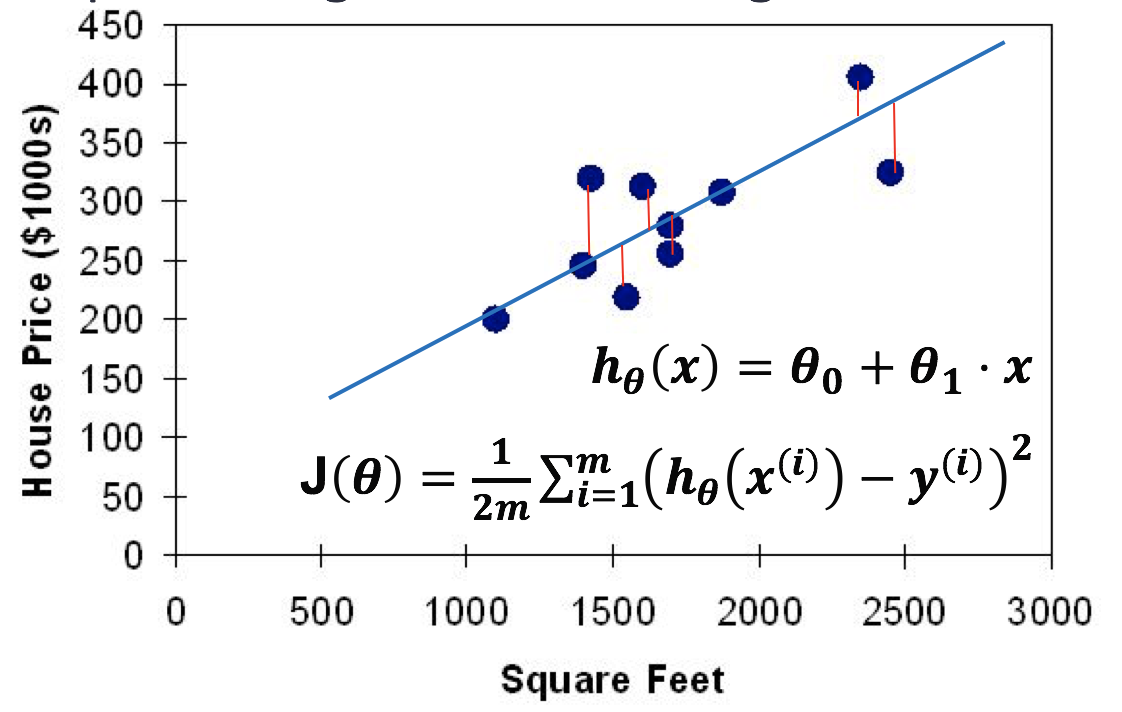
****

Using independent variables (features)

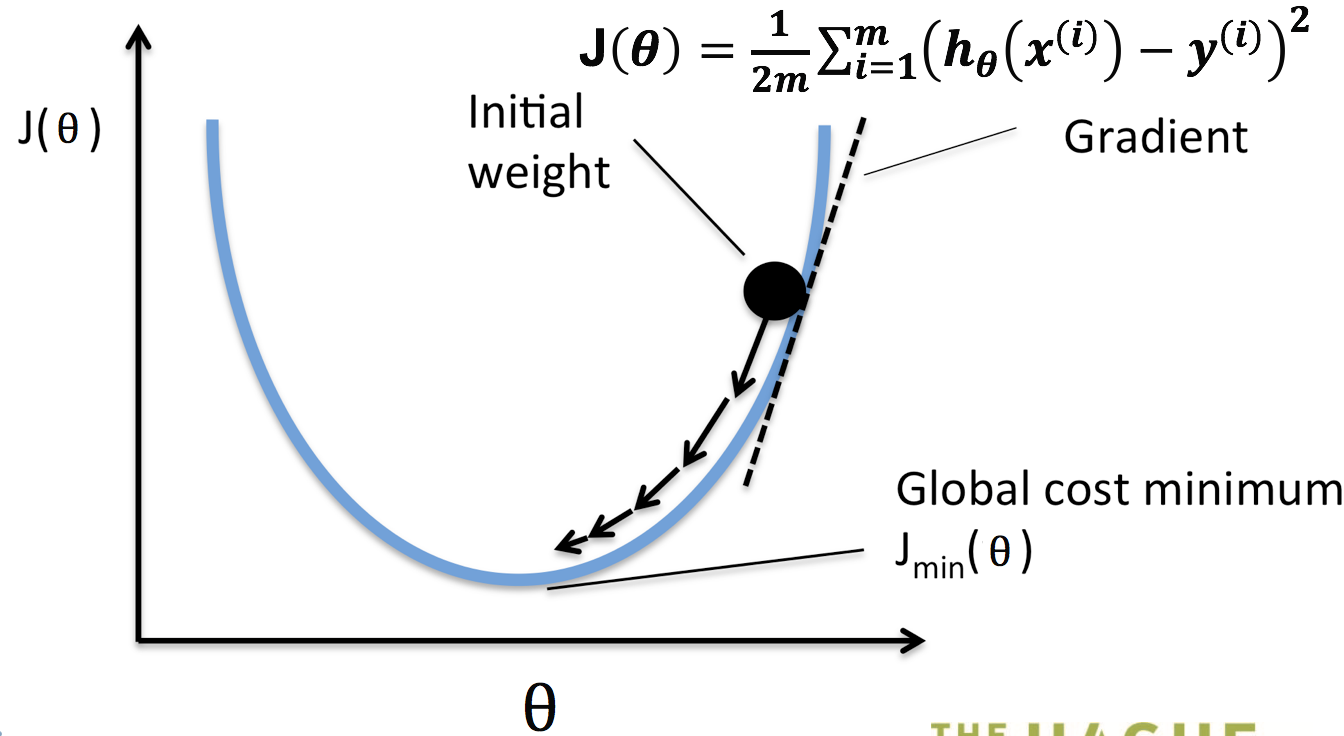
* Size
* Distance to city center

****

Find ‘optimal’ regression line using cost function



How to minimize a cost function



Update, by taking a small step towards minimum

