Supervised Learning: Regression

Parantapa Goswami

Viseo R&D
Grenoble, France
parantapa.goswami@viseo.com

December 1, 2017

Classification

IRIS Classification, Ronald Fisher (1936)

Classification

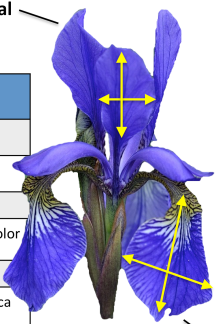
IRIS Classification, Ronald Fisher (1936)

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)



VISEO

Classification

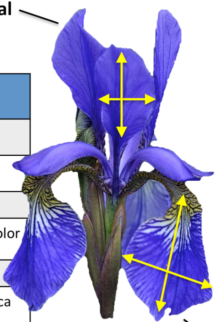
IRIS Classification, Ronald Fisher (1936)

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)



- discrete valued output

Classification

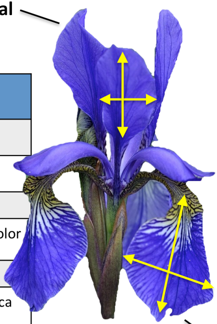
IRIS Classification, Ronald Fisher (1936)

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

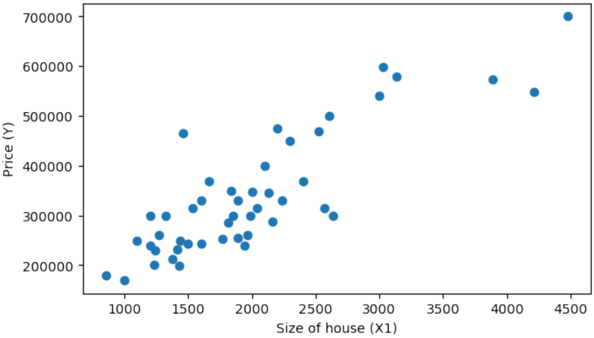
Features
(attributes, measurements, dimensions)

Class labels
(targets)



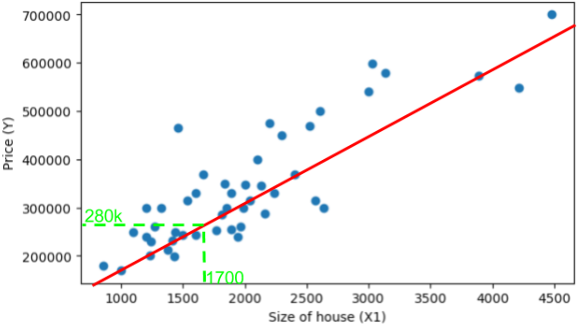
- discrete valued output
- what if output is continuous?

Regression



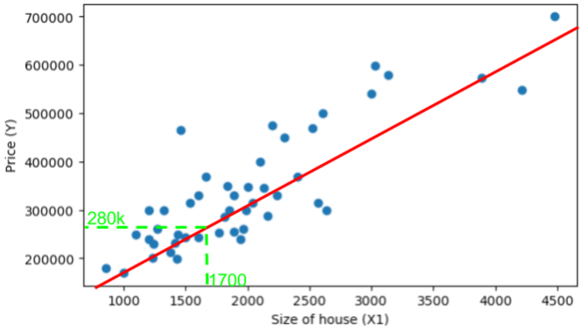
VISEO

Regression



700000
600000
500000
400000
300000
200000
Price (Y)
280k
1700
1000
1500
2000
2500
3000
3500
4000
4500
Size of house (X1)

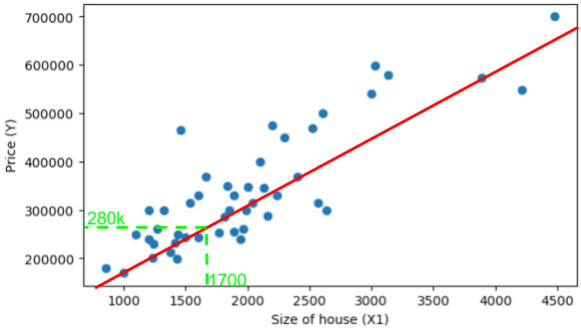
Regression



Supervised Learning

Right answers are given

Regression



Supervised Learning

Right answers are given

Regression Problem

Predict real valued output

Training Set

Size of house in feet^2 (x)	Price in 1000$s (y)
2104	460
1416	232
1534	315
⋮	⋮

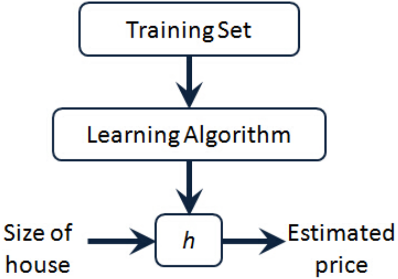
m → number of examples
x → input variable/features
y → output variable/target
(x^i, y^i) → i^{th} training example

VISEO

Training Set

Size of house in feet^2 (x)	Price in 1000$s (y)
2104	460
1416	232
1534	315
$\vdots$	$\vdots$

m $\rightarrow$ number of examples
x $\rightarrow$ input variable/features
y $\rightarrow$ output variable/target
(x^i, y^i) $\rightarrow$ i^{th} training example

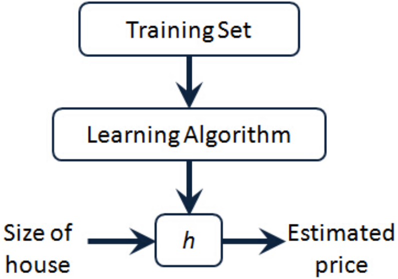


Hypothesis $h : x \rightarrow y$

Training Set

Size of house in feet^2 (x)	Price in 1000$s (y)
2104	460
1416	232
1534	315
⋮	⋮

m → number of examples
x → input variable/features
y → output variable/target
(x^i, y^i) → i^{th} training example

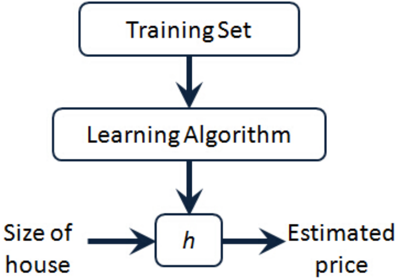


Hypothesis $h : x \to y$

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Training Set

Size of house in feet^2 (x)	Price in 1000$s (y)
2104	460
1416	232
1534	315
⋮	⋮



$m \rightarrow$ number of examples
$x \rightarrow$ input variable/features
$y \rightarrow$ output variable/target
$(x^i, y^i) \rightarrow i^{th}$ training example

Hypothesis $h : x \rightarrow y$

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Univariate Linear Regression

VISEO

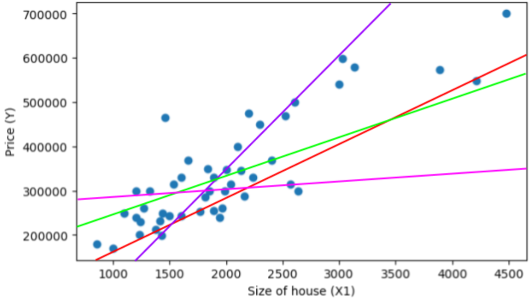
How to choose the parameters?

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

How to choose the parameters?

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

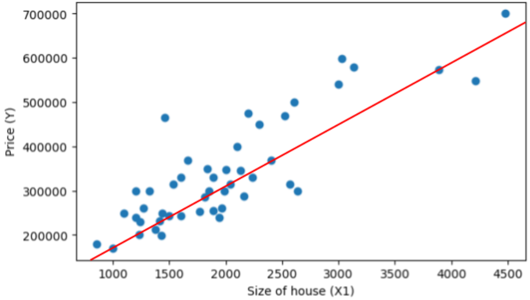
- Different θ_0 and θ_1 will generate different lines



How to choose the parameters?

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

- Different θ_0 and θ_1 will generate different lines



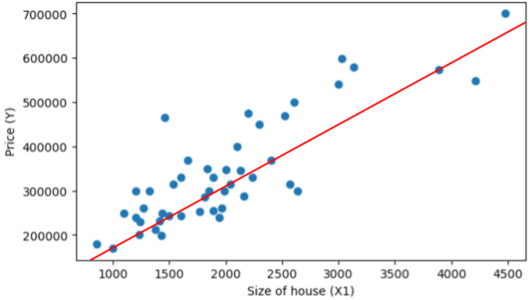
- Need to choose best θ_0 and θ_1

VISEO

How to choose the parameters?

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

- Different θ_0 and θ_1 will generate different lines



- Need to choose best θ_0 and θ_1

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Defining a Cost

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Defining a Cost

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Cost Function: Mean Squared Error

$$min_{\theta_0,\theta_1} \ (h_\theta(x) - y)^2$$

VISEO

Defining a Cost

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Cost Function: Mean Squared Error

$$min_{\theta_0,\theta_1} \ \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2$$

Defining a Cost

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Cost Function: Mean Squared Error

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

$$min_{\theta_0,\theta_1} \; J(\theta_0, \theta_1)$$

Defining a Cost

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Cost Function: Mean Squared Error

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

$$min_{\theta_0, \theta_1} \; J(\theta_0, \theta_1)$$

For fixed θ_0 and θ_1, h_θ is a function of x

VISEO

Defining a Cost

$$y = h_\theta(x) = \theta_0 + \theta_1 x$$

Idea: Choose θ_0 and θ_1 so that $h_\theta(x)$ is close to given outputs y in the training examples (x, y)

Cost Function: Mean Squared Error

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$$

$$min_{\theta_0, \theta_1} \; J(\theta_0, \theta_1)$$

For fixed θ_0 and θ_1, h_θ is a function of x

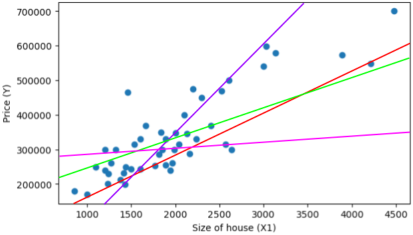
$J(\theta_0, \theta_1)$ is a function of the parameters θ_0 and θ_1

VISEO

More about the Cost

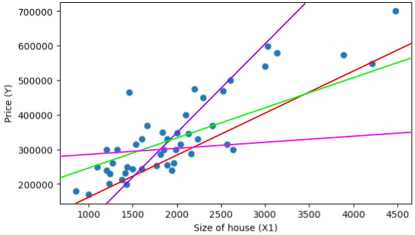
More about the Cost

For fixed θ_0 and θ_1, h_θ is a function of x

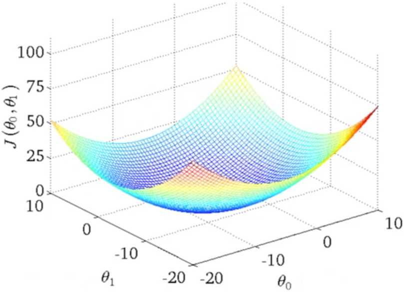


More about the Cost

For fixed θ_0 and θ_1, h_θ is a function of x



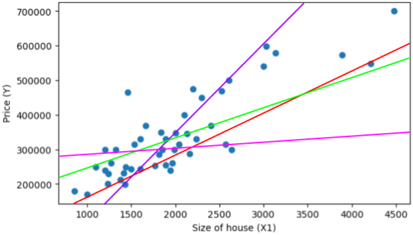
$J(\theta_0, \theta_1)$ is a function of the parameters θ_0 and θ_1

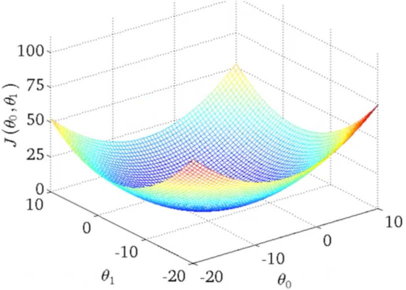


More about the Cost

For fixed θ_0 and θ_1, h_θ is a function of x

$J(\theta_0, \theta_1)$ is a function of the parameters θ_0 and θ_1





- Goal: to find a $\hat{\theta}_0$ and $\hat{\theta}_1$ providing minimum $J(\theta_0, \theta_1)$
- Corresponding hypothesis will be the trained model: $h_{\hat{\theta}}(x) = \hat{\theta}_0 + \hat{\theta}_1 x$
- $J(\theta_0, \theta_1)$ is convex: can be minimized using Gradient Descent

Multiple Features

Size of house in feet^2 (x)	Price in 1000$s (y)
2104	460
1416	232
1534	315
⋮	⋮

VISEO

Multiple Features

Size of house in feet^2 (x)	Number of bedrooms	Number of Floors	Price in 1000$s (y)
2104	5	1	460
1416	3	2	232
1534	3	2	315
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$n \rightarrow$ number of features
$\vec{x}^{i} \rightarrow$ i^{th} input is now a n-dimensional vector
$x_j^i \rightarrow$ j^{th} feature of the i^{th} input

Multivariate Linear Regression

Hypothesis

$$y = h_{\vec{\theta}}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Multivariate Linear Regression

Hypothesis

$$y = h_{\vec{\theta}}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Cost Function

$$J(\theta_0, \theta_1, \theta_2, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\vec{\theta}}(\vec{x^i}) - y^i)^2$$

VISEO

Multivariate Linear Regression

Hypothesis

$$y = h_{\vec{\theta}}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Cost Function

$$J(\theta_0, \theta_1, \theta_2, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_{\vec{\theta}}(\vec{x^i}) - y^i)^2$$

- Goal: to find $\hat{\theta}_0, \ldots, \hat{\theta}_n$ providing minimum $J(\theta_0, \ldots, \theta_n)$
- Here also, $J(\theta_0, \ldots, \theta_n)$ is convex: thus can be minimized using Gradient Descent

VISEO

Time Series

A time series is a sequence of observations $s_t \in \mathbb{R}$, usually ordered in time.

Examples of time series can be found in every scientific and applied domain:

- Meteorology: weather variables, like temperature, pressure, wind.
- Economy and finance: economic factors (GNP), financial indexes, exchange rate, spread.
- Marketing: activity of business, sales.
- Industry: electric load, power consumption, voltage, sensors.
- Bio-medicine: physiological signals (EEG), heart-rate, patient temperature.
- Web: clicks, logs.

VISEO

Time Series Forecasting

Time series forecasting is the use of a model to predict future values of a time series based on previously observed values.

Time Series Forecasting

Time series forecasting is the use of a model to predict future values of a time series based on previously observed values.

To do that:

1. Understand or model the stochastic mechanisms generating the data
2. This model is used to forecast the future values based on the history

VISEO

Time Series Forecasting

Time series forecasting is the use of a model to predict future values of a time series based on previously observed values.

To do that:

1. Understand or model the stochastic mechanisms generating the data
2. This model is used to forecast the future values based on the history

Examples:

- Weather forecasting.
- Sales prediction.
- Stock market forecasting.

Formalizing Time Series Forecasting

- Assume we have a time series:

$$x_1, x_2, \dots, x_N$$

Formalizing Time Series Forecasting

- Assume we have a time series:

$$x_1, x_2, \ldots, x_N$$

- T observed values: also known as history size

Formalizing Time Series Forecasting

- Assume we have a time series:

$$x_1, x_2, \ldots, x_N$$

- T observed values: also known as history size
- Need to forecast T' future values: also known as forecast horizon

Formalizing Time Series Forecasting

- Assume we have a time series:

$$x_1, x_2, \ldots, x_N$$

- T observed values: also known as history size
- Need to forecast T' future values: also known as forecast horizon
- We want to model the relation:

$$(x_{T+1}, x_{T+2}, \ldots, x_{T+T'}) = r(x_1, x_2, \ldots, x_T)$$

Constructing Regression Matrix

- To learn a regression model: multiple examples of feature value pairs are required

Constructing Regression Matrix

- To learn a regression model: multiple examples of feature value pairs are required
- Moving window method:

$$x_1, x_2, \ldots, x_T, x_{T+1}, x_{T+2}, \ldots x_{T+T'}, x_{T+T'+1}, \ldots$$

VISEO

Constructing Regression Matrix

- To learn a regression model: multiple examples of feature value pairs are required
- Moving window method:

$$\underbrace{x_1, x_2, \ldots, x_T}_{\text{history}}, \underbrace{x_{T+1}, x_{T+2}, \ldots, x_{T+T'}}_{\text{horizon}}, x_{T+T'+1}, \ldots$$

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_T \end{bmatrix} \quad Y = \begin{bmatrix} x_{T+1} & x_{T+2} & \ldots & x_{T+T'} \end{bmatrix}$$

VISEO

Constructing Regression Matrix

- To learn a regression model: multiple examples of feature value pairs are required
- Moving window method:

$$x_1, \underbrace{x_2, x_3, \ldots, x_T, x_{T+1}}_{\text{history}}, \underbrace{x_{T+2}, x_{T+3}, \ldots, x_{T+T'+1}}_{\text{horizon}}, x_{T+T'+2}, \ldots$$

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_T \\ x_2 & x_3 & \ldots & x_{T+1} \end{bmatrix} \quad Y = \begin{bmatrix} x_{T+1} & x_{T+2} & \ldots & x_{T+T'} \\ x_{T+2} & x_{T+3} & \ldots & x_{T+T'+1} \end{bmatrix}$$

Constructing Regression Matrix

- To learn a regression model: multiple examples of feature value pairs are required
- Moving window method:

$$x_1, x_2, \underbrace{x_3, \ldots, x_{T+1}, x_{T+2}}_{\text{history}}, \underbrace{x_{T+3}, x_{T+4}, \ldots, x_{T+T'+2}}_{\text{horizon}}, x_{T+T'+3}, \cdots$$

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_T \\ x_2 & x_3 & \ldots & x_{T+1} \\ x_3 & x_4 & \ldots & x_{T+2} \end{bmatrix} \quad Y = \begin{bmatrix} x_{T+1} & x_{T+2} & \ldots & x_{T+T'} \\ x_{T+2} & x_{T+3} & \ldots & x_{T+T'+1} \\ x_{T+3} & x_{T+4} & \ldots & x_{T+T'+2} \end{bmatrix}$$

Constructing Regression Matrix

- To learn a regression model: multiple examples of feature value pairs are required
- Moving window method:

$$x_1, x_2, \underbrace{x_3, \ldots, x_{T+1}, x_{T+2}}_{\text{history}}, \underbrace{x_{T+3}, x_{T+4}, \ldots, x_{T+T'+2}}_{\text{horizon}}, x_{T+T'+3}, \cdots$$

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_T \\ x_2 & x_3 & \ldots & x_{T+1} \\ x_3 & x_4 & \ldots & x_{T+2} \end{bmatrix} \quad Y = \begin{bmatrix} x_{T+1} & x_{T+2} & \ldots & x_{T+T'} \\ x_{T+2} & x_{T+3} & \ldots & x_{T+T'+1} \\ x_{T+3} & x_{T+4} & \ldots & x_{T+T'+2} \end{bmatrix}$$

- Regression models are learned on input X and output y

VISEO

A Good ML Course

- Introduction to Machine Learning. Course by Andrew Ng. (Youtube)

Hands On!

- 1_MachineLearning_Regression
- 2_TimeSeriesForecasting