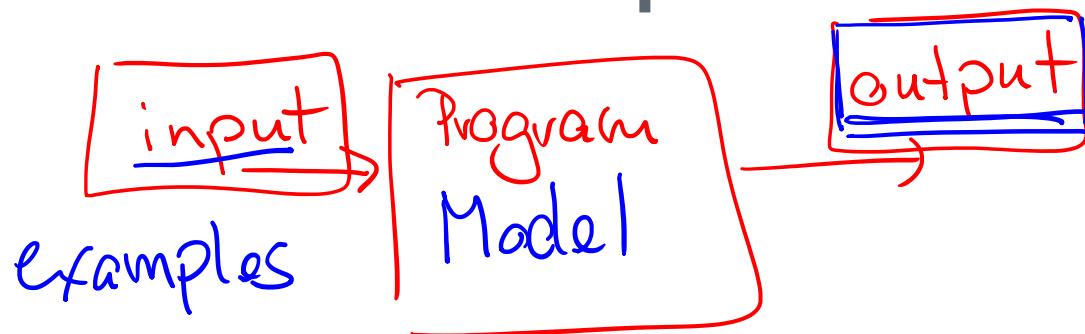

Introduction to Supervised Learning

Heni Ben Amor, Ph.D.
Assistant Professor
Arizona State University



Supervised Learning

- | Most common form of machine learning
- | Labeled training data of inputs and outputs
- | Labels often provided by an expert
- | Learns a mapping of input to output
- | Focus is on accurate predictions
- | Should also work on new inputs



Example: Diabetes Classification



- | Disease classification typically done by expert
- | Learning to automatically classify presence of Type 2 Diabetes

Example: Diabetes Classification

| Based on a set of readings: (*feature*)

- pregnancies: number of times pregnant
- glucose: plasma glucose concentration
- blood pressure: diastolic blood pressure
- skin thickness: triceps skin fold thickness (mm)
- Insulin: 2-hour serum insulin
- BMI: body mass index
- Diabetes pedigree: diabetes mellitus history in relatives
- Age: age in years

Example: Diabetes Classification

The diagram illustrates the structure of training and inference data for diabetes classification. It features two tables: a larger one for 'Training' data and a smaller one for 'Inference' data.

Training Data:

	Readings	Features	Label
2nd	0 109	88 30 0 32.5 0.855 38	1
3rd	2 109	92 0 0 42.7 0.845 54	0
:	1 95	66 13 38 19.6 0.334 25	0
	4 146	85 27 100 28.9 0.189 27	0
	2 100	66 20 90 32.9 0.867 28	1

Inference Data:

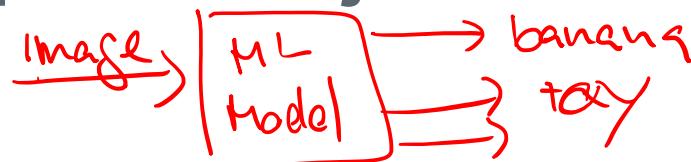
	Readings	Label
	5 139 64 35 140 28.6 0.411 26	?

Annotations in red and blue highlight specific elements:

- patient:** A red bracket on the left side of the first row of the training table.
- Training:** A red bracket above the first row of the training table.
- Readings:** A red bracket above the second column of the training table.
- Features:** A red bracket above the third column of the training table.
- Label:** A red bracket above the last column of the training table.
- 2nd:** A blue bracket next to the first row of the training table, indicating the second data point.
- 3rd:** A blue bracket next to the second row of the training table, indicating the third data point.
- Inference:** A blue bracket above the first row of the inference table.
- Readings:** A red bracket above the second column of the inference table.
- Label:** A red bracket above the last column of the inference table.
- 5:** A red circle around the value 5 in the first column of the inference table.

Example: Object Recognition

- | Recognition of objects in an image
- | Features are the pixel values (r, g, b) of the image
- | Classes are the types of objects to be detected



Generalization to Unseen Data

- | Learned model should generalize
- | Model applicable to wide range of variations
- | Holy Grail of ML: achieving high generalization



Types of Supervised Learning

| Classification:

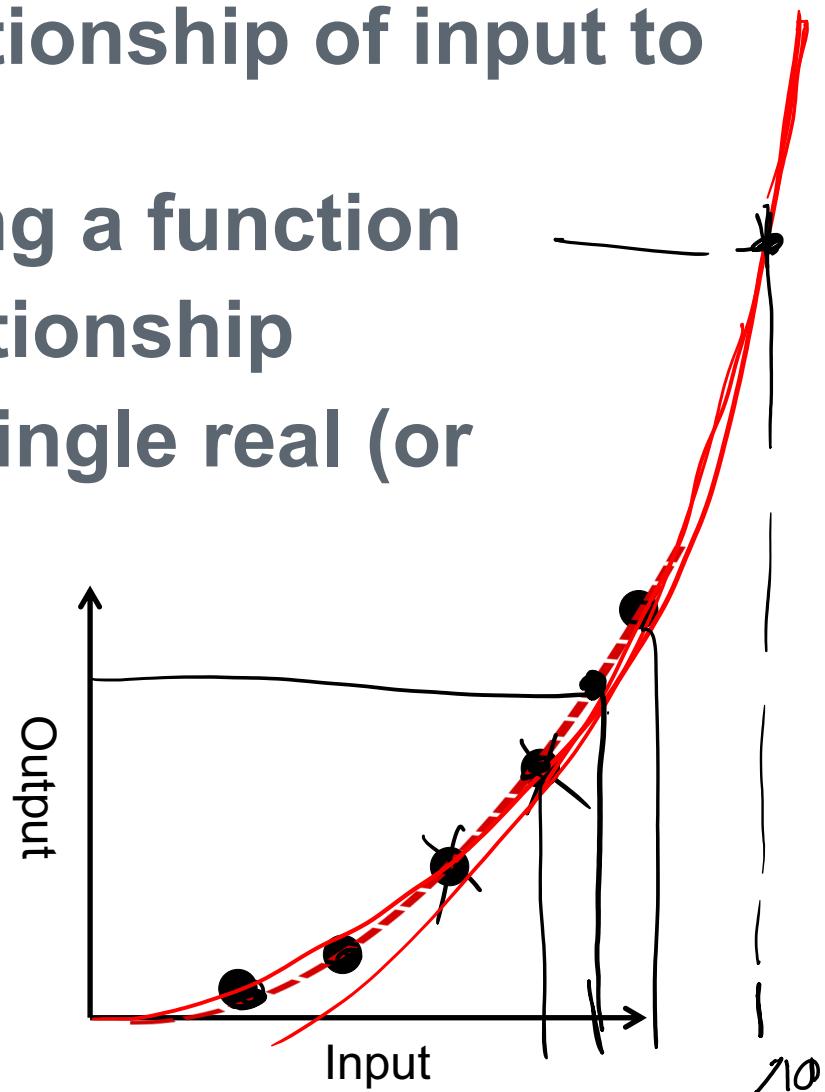
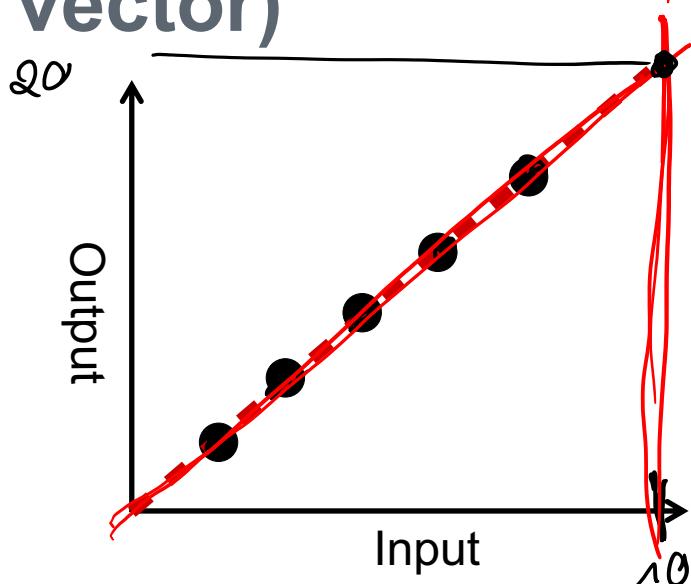
- The label defines a class to which the data point is assigned. We assume a discrete set of classes. Hence label is an integer denoting the number of the class.

| Regression:

- The label is one or multiple real numbers. Rather than discrete responses (classes), regression models continuous response values.

Regression

- | Learn continuous relationship of input to output
- | Can be seen as learning a function
- | Extract functional relationship
- | Maps input vector to single real (or vector)



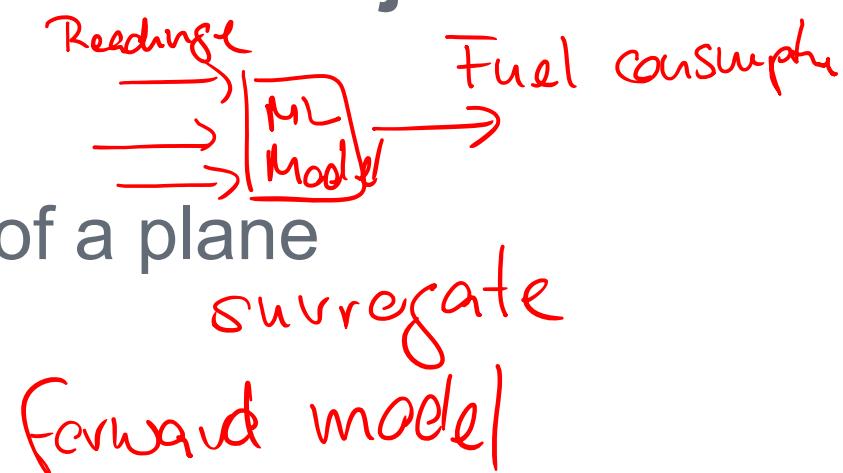
Example Applications of Regression

| Prediction and Forecasting

- Stock market predictions
- Predicting the output of a power plant
- Forecasting weather

| Creating Models of Real-world Objects

- Fuel consumption of car
- Airflow around the wing of a plane
- Movements of a robot



Loss Function



- | Often also called cost function or error function
- | Describes how well the given data is modeled
- | Measures the discrepancy to the target labels
- | High discrepancy yields a high loss

Loss Function

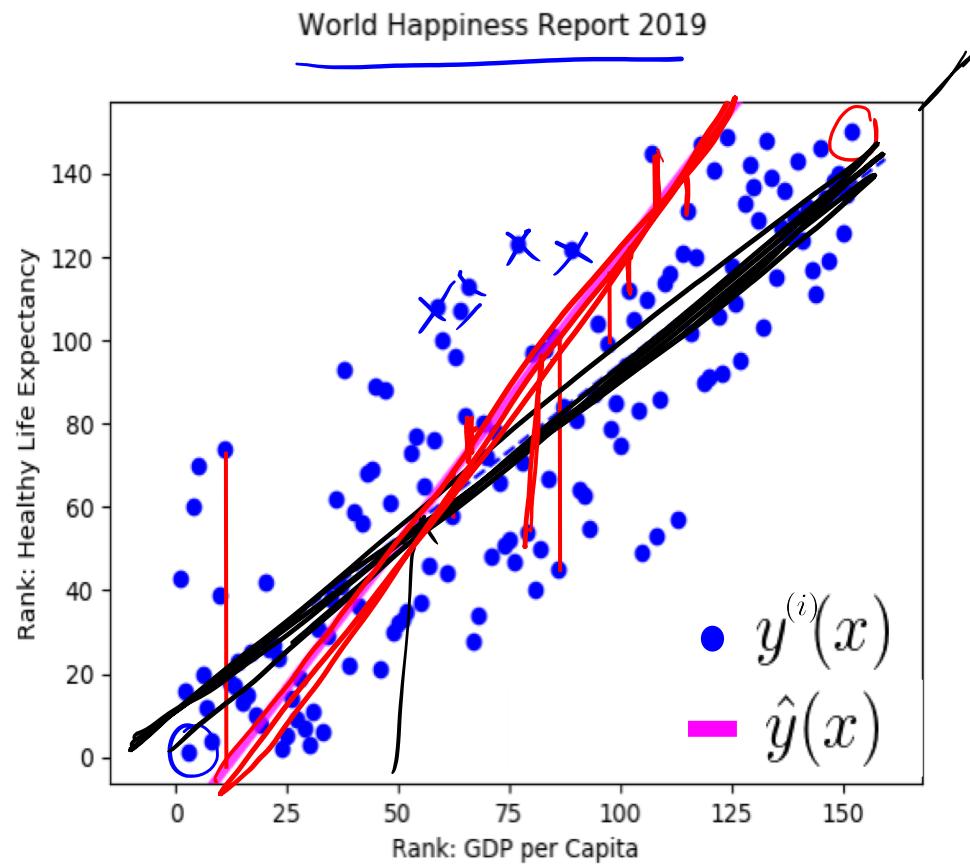


- | The goal is to find a model that minimizes loss.
- | Finding models that minimize the loss function is a central optimization problem at the core of most machine learning methods.

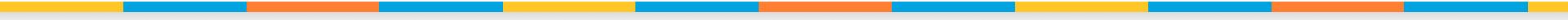
Example Loss Function

| Mean Square Error:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y^{(i)} - \hat{y}^{(i)})^2$$



Summary



- | **Introduced supervised learning**
- | **Most common type of machine learning**
- | **Learn a relationship from inputs to outputs**
- | **Labeled training data is provided by expert**
- | **Two variants**
 - Classification
 - Regression
- | **Loss function plays a central role**