We are given a **training set** which consists of training examples .

Place the entire training set into a table as follows.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample Number  (Training Example Number) | feature 1 | feature 2 |  | feature | () target |
| 1 |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Feature Vector:

* One row of features (stored as a column vector)

Target Value:

Target Vector:

Training Example:

**Design Matrix**:

* Feature vectors as rows

**MSE Cost Function**:

**Closed Form Least Squares Solution for Linear Regression**

**Regularized Least Squares and MAP:**

Recall

* problematic if poorly conditioned
* Solve by adding to diagonal.0

**Ridge Regression**

RR is solution to

* The term cause us to fit the line.
* The term penalize the magnitude of the parameters.

Derivation

**Cross Validation**

* Compute the train error and the dev-error for different regularization factors , and choose record the maximum of the two errors for each value of .
* Then, choose the minimum of these maximum values (min-of-max).

**Logistic Regression:**

Input (Data) Matrix:

Output (Target) Vector:

Given produce an estimate of the target, donated .

Parameters of model

Logistic function:

Hypothesis:

**Loss function** (1-training example):

* Squared error:
* Cross-Entropy:
  + Theoretical underpinnings:
    - We want output as the probability that given the data
    - Another way of saying this is:
    - Combining:
* If , then
* If , then

**Logistic Regression Backprop**:

**Backprop** with network with training examples.

**Backprop** of network for binary classification ( output neurons) and training examples with:

* Loss Function:
  + Cross Entropy
* Activation Functions:
  + Sigmoid

Output-layer changes as follows:

**MINUS SIGN IS OFF**

**Backprop** of network for binary classification ( output neurons) and training examples with:

* Loss Function:
  + Cross Entropy
* Activation Functions:
  + ReLu

**Softmax**:

Consider modifying logistic regression to have three neurons followed by a Softmax.

Consider

Note:

**Indicator function**

**Loss**

**Cost**

**Backprop** of network for multi-class classification ( output neurons) and training examples with:

* Loss Function:
  + Cross Entropy
* Output activation Functions:
  + Softmax

**CNN Forward Prop**

Single Example