**Multivariate Linear Regression**

|  |
| --- |
| **Housing price – multi-features example** |
|  |
| **Notations** |
|  |
| **Multivariate Hypothesis** |
| n features:  Note:      Therefore, |

**Multivariant Gradient Descent**

|  |
| --- |
| **Summarize** |
| A n+1-dimensioned vector  **Equivalent Cost Function Formulas** |
| **Multivariate Gradient Descent** |
| Simultaneous update for j = [0…n] ()  **Multivariate Simultaneous Update (n >= 1)**  … |

**Tips for Optimizing Gradient Descent**

|  |
| --- |
| **Purpose: Make gradient decent run faster with smaller data** |
| 1. **Feature Scaling** |
| **Idea** |
| Make sure features are on a similar scale |
| **Significance** |
| Such that more directed path to local minimum can be found |
| **Goal** |
| Get every feature / variable into approximately -1 <= x (i) <= 1 range |
| **Example** |
| |  |  |  |  | | --- | --- | --- | --- | | **Before Scaling** | |  | **After Scaling** | |  | 🡪 | |  | |
| 1. **Mean Normalization** |
|  |
| **Goal: To chose a learning rate α, make sure gradient descent working correctly** |
| Try …0.001, 0.01, 0.1, 1… |

**Features and Polynomial Regression**

|  |
| --- |
| **Square root function with 1 feature: size** |
| **Feature Scaling** |
| **Cubic function with 1 feature: size** |
| **Feature scaling**  If x1 has range 1-1000, then x3 has range 1-10^9. |
| **Combine multi feature into 1** |
| For example: 2 variable: width and length can be combine into size = width \* length |

**Computing Parameters Analytically**

|  |
| --- |
| **Normal Equation** |
| Method to solve optimum local minimum in 1 step |
| **Example** |
|  |
|  |
| **Gradient Descent vs. Normal Equation** |
| |  |  |  | | --- | --- | --- | |  | Gradient Descent | Normal Equation | | Advantages and disadvantage | Need to choose α  Needs many iterations  Works well when n is large | Don’t need to choose α  Don’t need to iterate  Slow if n is large  Need to compute (XT\*X)-1 O(n^3) | | When to use | When n is large n > 1000 | When n is small n < 1000 | |
|  |
| **Normal Equation Non-invertibility** |
| |  |  | | --- | --- | |  | **Cause of non-invertible XT\*X** | | 1 | **Redundant Feature** | |  | If there is a linear relationship between two variables x2 = 3.28\*x1 | | 2 | **Too many features** | |  | When #features > #training sample **(m <= n)** | |