Machine Learning

Announcements

Prediction

Regression

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

Regression

```
Given a set of (x, y) pairs, find a function f(x) that returns good y values

   pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Data from home sales records in Ames, Iowa

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Square feet

Data from home sales records in Ames, Iowa

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Square feet

Price (thousands)

Data from home sales records in Ames, Iowa

Regression

`Given a set of (x, y) pairs, find a function f(x) that returns good y values`

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Square feet

Price (thousands)

Data from home sales records in Ames, Iowa

Measuring error: $|y-f(x)|$ or $(y-f(x))^2$ are both typical

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Square feet

Price (thousands)

Data from home sales records in Ames, Iowa

Measuring error: $|y-f(x)|$ or $(y-f(x))^2$ are both typical

Over the whole set of (x, y) pairs, we can compute the mean of the squared error

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Square feet

Price (thousands)

Data from home sales records in Ames, Iowa

Measuring error: |y−f(x)| or $(y-f(x))^2$ are both typical

Over the whole set of (x, y) pairs, we can compute the mean of the squared error

Squared error has the wrong units, so it's common to take the square root

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Square feet

Price (thousands)

Data from home sales records in Ames, Iowa

Measuring error: $|y-f(x)|$ or $(y-f(x))^2$ are both typical

Over the whole set of (x, y) pairs, we can compute the mean of the squared error

Squared error has the wrong units, so it's common to take the square root

The result is the "root mean squared error" of a predictor f on a set of (x, y) pairs

Regression

Given a set of (x, y) pairs, find a function f(x) that returns good y values

```
pairs = [(1656, 215.0), (896, 105.0), (1329, 172.0), ...]
```

Data from home sales records in Ames, Iowa

Square feet

Price (thousands)

Measuring error: $|y-f(x)|$ or $(y-f(x))^2$ are both typical

Over the whole set of (x, y) pairs, we can compute the mean of the squared error

Squared error has the wrong units, so it's common to take the square root

The result is the "root mean squared error" of a predictor f on a set of (x, y) pairs

(Demo)

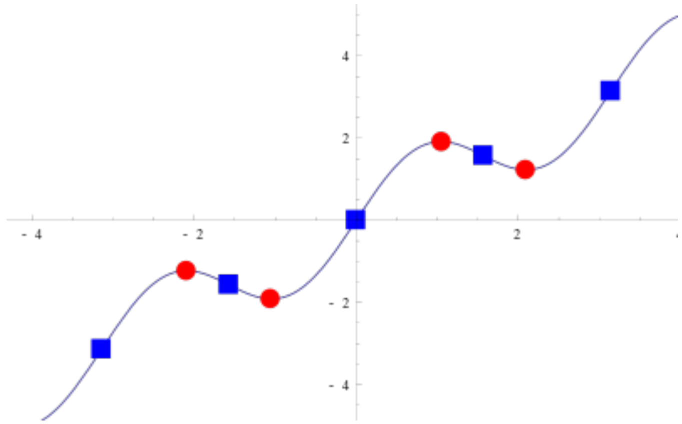
Critical Points

Critical Points

```
Maxima, minima, and inflection points of a differentiable function occur
when the derivative is 0
```

Critical Points

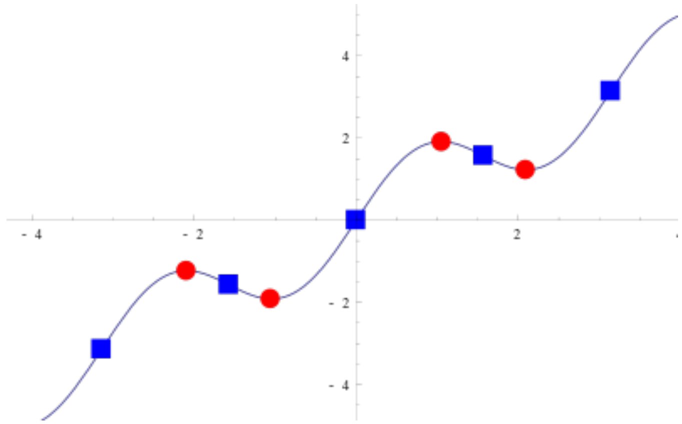
Maxima, minima, and inflection points of a differentiable function occur when the derivative is 0



http://upload.wikimedia.org/wikipedia/commons/f/fd/Stationary_vs_inflection_pts.svg

Critical Points

Maxima, minima, and inflection points of a differentiable function occur when the derivative is 0

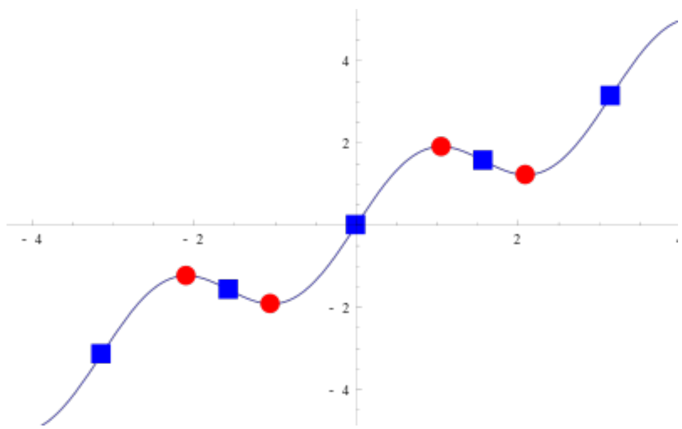


The global minimum of convex functions that are (mostly) twice-differentiable can be computed numerically using techniques that are similar to Newton's method

http://upload.wikimedia.org/wikipedia/commons/f/fd/Stationary_vs_inflection_pts.svg

Critical Points

Maxima, minima, and inflection points of a differentiable function occur when the derivative is 0



The global minimum of convex functions that are (mostly) twice-differentiable can be computed numerically using techniques that are similar to Newton's method

(Demo)

http://upload.wikimedia.org/wikipedia/commons/f/fd/Stationary_vs_inflection_pts.svg

Multiple Linear Regression

```
Given a set of (xs, y) pairs, find a linear function f(xs) that returns good y values
```

Multiple Linear Regression

Given a set of (xs, y) pairs, find a linear function f(xs) that returns good y values

A linear function has the form **w** • **xs** + *b* for vectors **w** and **xs** and scalar *b*

Multiple Linear Regression

Given a set of (xs, y) pairs, find a linear function f(xs) that returns good y values

A linear function has the form **w • xs** + *b* for vectors **w** and **xs** and scalar *b*

(Demo)

Multiple Linear Regression

Given a set of (xs, y) pairs, find a linear function f(xs) that returns good y values

A linear function has the form **w • xs** + *b* for vectors **w** and **xs** and scalar *b*

(Demo)

Note: Root mean squared error can be optimized through linear algebra alone, but numerical optimization works for a much larger class of related error measures