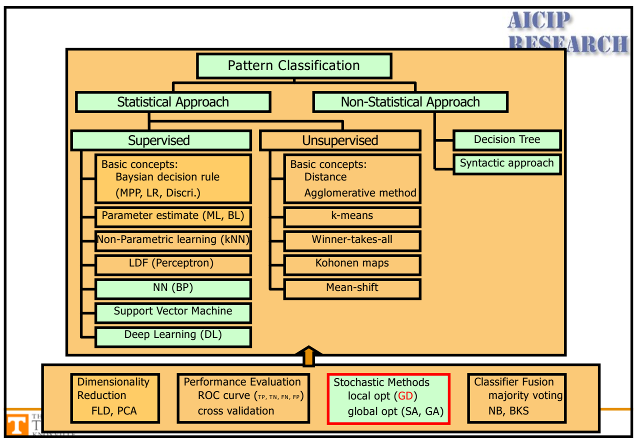
THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

ECE471-571 – Pattern Recognition

Lecture 14 – Gradient Descent

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu



AICIP RESEARCH

Review - Bayes Decision Rule

$$P(\omega_j | x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

Maximum Posterior Probability: For a given *x*, if $P(\omega_1|x) > P(\omega_2|x)$, then *x* belongs to class 1, otherwise 2

Likelihood Ratio: If $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$, then *x* belongs to class 1, otherwise, 2.

Discriminant Function: The classifier will assign a feature vector x to class ω_i if $g_i(x) > g_j(x)$

- Case 1: Minimum Euclidean Distance (Linear Machine), $\Sigma_i = \sigma^2 I$
- Case 2: Minimum Mahalanobis Distance (Linear Machine), $\Sigma_i = \Sigma$
- Case 3: Quadratic classifier, Σ_i = arbitrary

Non-parametric kNN: For a given *x*, if $k_1/k > k_2/k$, then *x* belongs to class 1, otherwise 2

Estimate Gaussian (Maximum Likelihood Estimation, MLE), Two-modal Gaussian

Dimensionality reduction

Performance evaluation ROC curve

THE UNIVERSITY OF TENNESSEE

3

General Approach to Learning

AICIP RESEARCH

- Optimization methods
  - Newton's method
  - Gradient descent
  - Exhaustive search through the solution space
- Objective functions
  - Maximum a-posteriori probability
  - Maximum likelihood estimate
  - Fisher's linear discriminant
  - Principal component analysis
  - k-nearest neighbor
  - Perceptron

THE UNIVERSITY OF TENNESSEE KNOXVILLE 4

General Approach to Learning

AICIP RESEARCH

- Specify a model (objective function) and estimate its parameters
- Use optimization methods to find the parameters
  - 1st derivative = 0
  - Gradient descent
  - Exhaustive search through the solution space

THE UNIVERSITY OF TENNESSEE KNOXVILLE 5

Newton-Raphson Method

AICIP RESEARCH

Used to find solution to equations

According to Taylor series: $f(x+\Delta x) \approx f(x) + \Delta x f'(x)$

$$f(x) + \Delta x f'(x) = 0 \Rightarrow \Delta x = -\frac{f(x)}{f'(x)}$$

$$\Rightarrow x^{k+1} = x^k - \frac{f(x)}{f'(x)}$$

THE UNIVERSITY OF TENNESSEE KNOXVILLE 6

Newton-Raphson Method vs. Gradient Descent

AICIP RESEARCH

f(x) = $x^2 - 5x - 4$
f(x) = xcosx

- Newton-Raphson method
- Used to find solution to equations
  - Find x for f(x) = 0
- The approach
  - Step 1: select initial x0
  - Step 2:

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

  - Step 3: if $|x^{k+1} - x^k| < \varepsilon$, then stop; else $x^k = x^{k+1}$ and go back step 2.

- Gradient descent
- Used to find optima, i.e. solutions to derivatives
  - Find x* such that f(x*) < f(x)
- The approach
  - Step 1: select initial x0
  - Step 2:

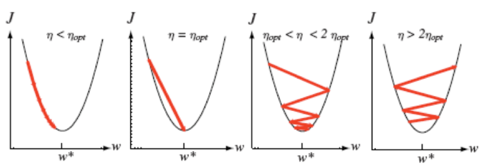
$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} = x^k - cf'(x^k)$$

  - Step 3: if $|x^{k+1} - x^k| < \varepsilon$, then stop; else $x^k = x^{k+1}$ and go back step 2.

On the Learning Rate

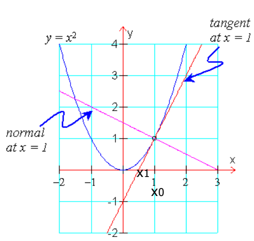
AICIP RESEARCH

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)} = x^k - cf'(x^k)$$



Geometric Interpretation

AICIP RESEARCH



Gradient of tangent is 2

http://www.teacherschoice.com.au/Maths_Library/Calculus/tangents_and_normals.htm

AICIP RESEARCH

```
% [x] = gd(x0, c, epsilon)
%   - Demonstrate gradient descent
%   - x0: the initial x
%   - c: the learning rate
%   - epsilon: controls the accuracy of the solution
%   - x: an array tracks x in each iteration
%
% Note: try c = 0.1, 0.2, 1; epsilon = 0.01, 0.1
function [x] = gd(x0, c, epsilon)

% plot the curve
clf;    fx = -10:0.1:10;    [fy, dfy] = g(fx);    plot(fx, fy);    hold on;

% find the minimum
x(1) = x0;    finish = 0;    i = 1;
while ~finish
   [y, dy] = g(x(i));

   plot(x(i), y, 'r*');    pause
   x(i+1) = x(i) - c * dy;
   if abs(x(i+1)-x(i)) < epsilon
      finish = 1;
   end
   i = i + 1;
end
x(i)

% plot trial points
[y, dy] = g(x);
for i=1:length(x)
   plot(x(i), y(i), 'r*');
end
```

```
function [y, dy] = g(x)
y = 50*sin(x) + x .* x ;
dy = 50*cos(x) + 2*x;
```

10