~~with L2 regularizer~~

$$\cancel{L' = L + C\sum_c w_c^T w_c}$$

~~if the derivative is over the same index (diag) this adds 2C, making it +2C~~

II Newton's Method (1-D)

Taylor Series → approx w/ Quadratic

$$f(\hat{x}) + f'(\hat{x})(x-\hat{x}) + \frac{f''(\hat{x})}{2}(x-\hat{x})^2$$

SS

$$y = ax^2 + bx + c$$

$$f'(x_i) = 2ax_i + b$$

$$f''(x_i) = 2a \Longrightarrow a = \frac{f''(x_i)}{2}$$

$$f'(x_i) = f''(x_i)x + b \Longrightarrow b = f'(x_i) - f''(x_i)x_i$$

$c \Rightarrow$ who cares?

Minimum $\overset{x_{i+1}}{@}$ $2ax_{i+1}+b=0$

$$f''(x_i)x_{i+1}+f'(x_i)-f''(x_i)x_i=0$$

$$x_j = x_i - \frac{f'(x_i)}{f''(x_i)}$$

DEMO

III Newton / Quasi-Newton

$$X_{n+1} = X_n - H\nabla$$

BFGS - set $H_0 = I$ $x_0 = ?$

$y_k = \nabla(x_{k+1}) + \nabla(x_k)$ $H_{k+1} = f(H_k, y_k)$

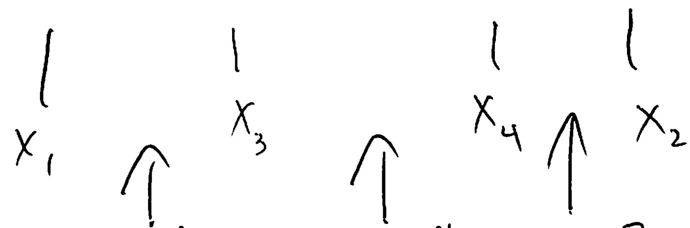
if H_k +Def $\Rightarrow$ H_{k+1} is +Def

L-BFGS

Compute H on the fly each time from the last m gradients.

Line Search

Brackets



Where is the min?

Repeat until wolfe conditions are satisfied

1. objective has decreased "enough"
2. |slope| has decreased "enough"

defined by constants

IV In Practice

Matlab

show example file

Java - Mallet

show example file