wFall2020 lecture4

Tuesday, September 22, 2020 9:18 PM

- homework 0
- homework 1
- honor code

Rewiew:

- lineer regression
  - loss: mean squared loss
    motivation: gaussian, max likelihood
  - optimization: GD, SGD.
- logistic regression
  - probabilistic model.
    $$p(y=1 \mid x;\theta) = h_\theta(x) = g(\theta^T x)$$
  - loss: max likelihood
  - optimization: GD, SGD.

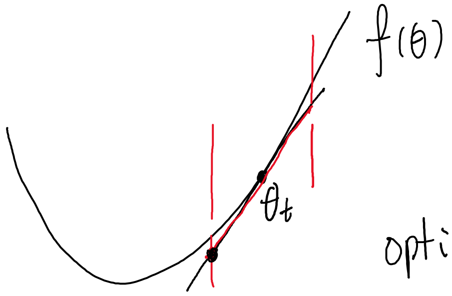
Overview:

1. Newton's method.
2. ML pipleline / hygiene.
3. generalized linear model.
   - exponential family

Newton's method

GD : another interpretation

$f'(\theta_t)(\theta-\theta_t)$

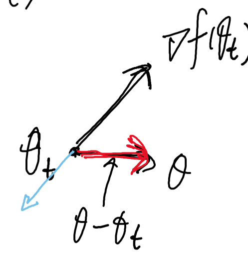


$$f(\theta) \approx \underline{f(\theta_t) + \langle \nabla f(\theta_t), \theta-\theta_t\rangle}$$

optimize local approx.

$$\arg\min_\theta \; \cancel{f(\theta_t)} + \langle \nabla f(\theta_t), \theta-\theta_t\rangle$$

s.t. $\|\theta-\theta_t\|_2 \le \varepsilon$

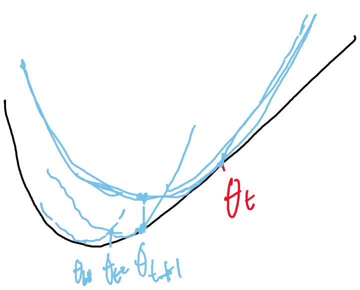


$$= -\alpha \nabla f(\theta_t)$$

$\alpha > 0$ scalar

s.t. $\|\alpha \nabla f(\theta_t)\| = \varepsilon$

Newton's



$$f(\theta) \approx \overbrace{f(\theta_t) + \langle \nabla f(\theta_t), \theta-\theta_t\rangle + \frac{1}{2}\langle \theta-\theta_t, \nabla^2 f}^{\tilde{f}_t(\theta)}$$

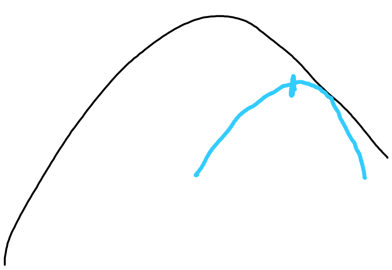
$$\boxed{f(\theta_t) + f'(\theta_t)(\theta-\theta_t) + \frac{1}{2} f''(\theta_t)(\theta-\theta_t}$$

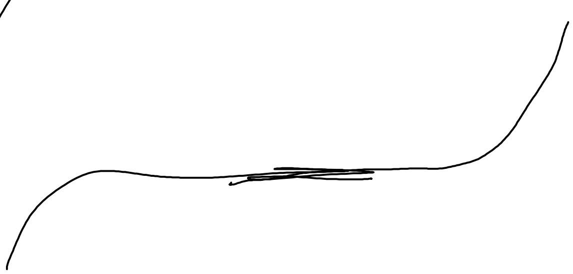
$$\boxed{\theta_{t+1} = \arg\min_\theta \tilde{f}_t(\theta)}$$

$$\nabla \tilde{f}_t(\theta) = 0$$

$$\nabla \tilde{f}_t(\theta) = \nabla f(\theta_t) + \nabla^2 f(\theta_t)(\theta-\theta_t) =$$

$$\boxed{f'(\theta_t) + f''(\theta_t)(\theta-\theta_t)}$$





$$\theta - \theta_t = -\nabla^2 f(\theta_t)^{-1} \cdot \nabla f(\theta_t)$$

$$\theta = \theta_t - \boxed{\nabla^2 f(\theta_t)^{-1}} \nabla f(\theta_t)$$

ML hygiene

Given dataset $\{(x, y) \text{ pairs}\}$ | train the parame

Split data randomly into
- training set $\{\{x^{(1)}, y^{(1)}\}, \ldots, \{x^{(n)},$
- development / validation set
- test set (blind set)

training stage:
- define probabalistic model parametrized by t
- derive loss func

}