Online Regression via O.G.D.

Logistical comments:

- Project instructions are now on website
- Components of project:
  * Proposal
  * ~~Short Presentation~~
  * Report
- Teams of 1–3

I Survey-type

II Original Research

I+II Combo

Proposal max 1 page
- Team members
- Topic
- Initial survey

Due date: end of next weekend

Please see website for more!

Online regression: (LINEAR)

x_t context

y_t response

Q How is response generated?

I "Realizable"

$$y_t = \langle w^*, x_t \rangle$$

fixed unknown

II "Stochastic response"

$$y_t = \langle w^*, x_t \rangle + \xi_t$$

mean-zero noise

III Arbitrary response

- y_t is chosen by Nature
- Compete w/ best linear predictor w^*_T

$$w^* = \arg\min_w \sum_{t=1} (y_t - \langle w, x_t \rangle)^2$$

Technical assumptions:

$|y_t| \le B$

$\|x_t\| \le R$

Goal: minimize ~~E~~ regret

I $y_t = \langle w^*, x_t \rangle$

$$\text{Regret} = \frac{1}{T}\sum_{t=1}^{T} \left(\langle w^* - w_t, x_t \rangle \right)^2$$

Algorithm ?

Gradient Descent

Online

$$\ell(y, \hat{y}) = (y - \hat{y})^2$$

$$\frac{\partial}{\partial \hat{y}} \ell(y, \hat{y}) = -2(y - \hat{y})$$

$$= 2(\hat{y} - y)$$

$$\hat{y}_t = \langle w_t, x_t \rangle$$

$$\nabla_w \ell(y_t, \langle w, x_t \rangle)$$

$$= 2(\langle w, x_t \rangle - y) x_t$$

w_{t-1}

$$w_t = w_{t-1} - \underbrace{\eta}_{\text{step size}} \nabla_{w_{t-1}} \ell(y_t, \langle w, x_t \rangle)$$

$$= w_{t-1} - 2\eta(\langle w_{t-1}, x_t \rangle - \underbrace{y_t}_{\langle w^*, x_t \rangle}) x_t$$

$$\|w_t - w^*\|^2$$

$$= \|w_{t-1} - w^* - 2\eta \langle w_{t-1} - w^*, x_t \rangle \cdot x_t\|^2$$

$$\|w^* - w_{t-1} - 2\eta\langle w_{t-1}-w^*, x_t\rangle x_t\|^2$$

$$= \|w^*-w_{t-1}\|^2 + \|x_t\|^2 4\eta^2 \langle w_{t-1}-w^*, x_t\rangle^2 - 4\eta\langle w_{t-1}-w^*, x_t\rangle^2$$

$$\Downarrow$$

$$0 \le \|w^*-w_{T+1}\|^2$$

$$= \|w^*-w_0\|^2 + \sum_{t=1}^{T} 4\eta^2\|x_t\|^2\langle w_{t-1}-w^*, x_t\rangle^2 - \sum_{t=1}^{T} 4\eta\langle w_{t-1}-w^*, x_t\rangle^2$$

$$\sum_{t=1}^{T+1} 4\eta^2 \underbrace{\|x_t\|^2}_{\le R^2}\langle w_{t-1}-w^*, x_t\rangle^2$$

$$\le 4\eta^2R^2(\text{Regret})$$

$$= -4\eta(\text{Regret})$$

$$0 \le \|w^*-w_0\|^2 + 4\eta^2R^2(\text{Regret}) - 4\eta(\text{Regret})$$

$$\eta = \frac{1}{2R^2}$$

$$= \|w^*-w_0\|^2 - 2\eta(\text{Regret})$$

$$2\eta(\text{Regret}) \leq \|w^* - w_0\|^2$$

$$\eta = \frac{1}{2R^2}$$

$$(\text{Regret}) \leq R^2 \|w^* - w_0\|^2$$

$$\parallel$$

$$\sum_{t=1}^{T} \langle w^* - w_{t-1}, x_t \rangle^2$$

Stochastic case:

$$y_t = \langle w^*, x_t \rangle + \xi_t, \quad w_t = w_{t-1} - 2\eta(\langle w_{t-1}, x_t \rangle - y_t) x_t$$

$$= w_{t-1} - 2\eta(\langle w_{t-1} - w^*, x_t \rangle - \xi_t) x_t$$

$$\|w_t - w^*\|^2$$

$$= \|w_{t-1} - w^*\|^2 + 4\eta^2 \|x_t\|^2 (\langle w_{t-1} - w^*, x_t \rangle - \xi_t)^2$$

$$- 2\eta(\langle w_{t-1} - w^*, x_t \rangle - \xi_t)$$

$$\langle w_{t-1} - w^*, x_t \rangle$$

Inequality: $(a+b)^2 \le 2a^2 + 2b^2$

$$(\langle w_{t-1} - w^*, x_t \rangle - \xi_t)^2 \le 2\langle w_{t-1} - w^*, x_t \rangle^2 + \xi_t^2$$

Assume: $\mathbb{E}\,\xi_t^2 = \sigma^2$

$$\mathbb{E}\|w_t - w^*\|^2 \le \mathbb{E}\|w_{t-1} - w^*\|^2 + 8\eta^2 R^2 \left(\mathbb{E}\langle w^* - w_{t-1}, x_t\rangle^2 + \mathbb{E}\,\xi_t^2\right) - 4\eta\,\mathbb{E}\langle w^* - w_{t-1}, x_t\rangle^2$$

$$0 \le \mathbb{E}\|w_T - w^*\|^2 \le \|w_0 - w^*\|^2 + 8\eta^2 R^2 (\mathbb{E}\,\text{Regret}) + 8\eta^2 R^2 \sum_{t=1}^{T} \mathbb{E}\,\xi_t^2 \left(\le 8\eta^2 R^2 T \sigma^2\right) - 4\eta(\mathbb{E}\,\text{Regret})$$

$$\eta = \min\left\{\frac{1}{4R^2}, \frac{1}{\sigma R\sqrt{T}}\right\}$$

$$\le \|w_0 - w^*\|^2 + 8 - 2\eta(\mathbb{E}\,\text{Regret})$$

$$\mathbb{E}\,\mathrm{Regret} \leq \frac{1}{\eta}\left(8 + \|w^* - w_0\|^2\right)$$

$$\leq \max\{\sigma R\sqrt{T}, 4R^2\}\left(8 + \|w^* - w_0\|^2\right)$$