An Analysis of the Chaudhuri and Monteleoni Algorithm

--- a differentially private logistic regression algorithm

Cynthia Dwork, Parikshit Gopalan, Huijia (Rachel) Lin,
Toniann Pitassi, Guy Rothblum, Adam Smith, Sergey Yekhanin

Logistic Regression

- Used to predict the probability of an event
  by learning weights on different attributes. ?

Age	Female	Single	Clicked	Salary
40	1	1	1	10K

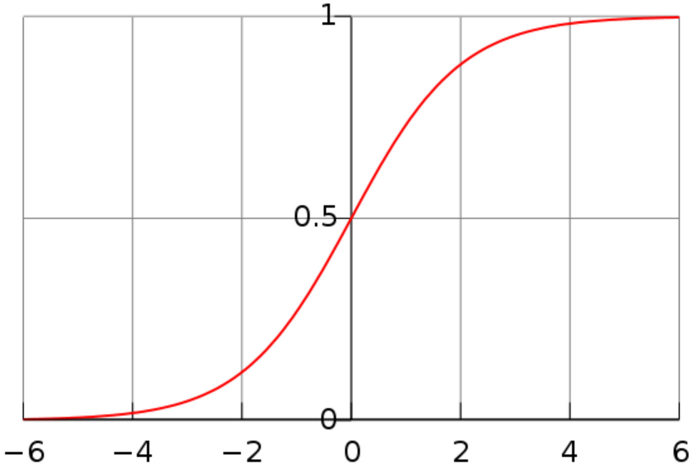
Weights:

.03	.5	.05	.001	1.5

Weighted Sum:
z = 40*0.03 + 1*0.5 + 1*0.05
1*.001 + 10*1.5
= 27.551

$$P = \frac{1}{1+e^{-z}}$$

Logistic Function



Logistic Regression

Fit the training data to Logistic Func

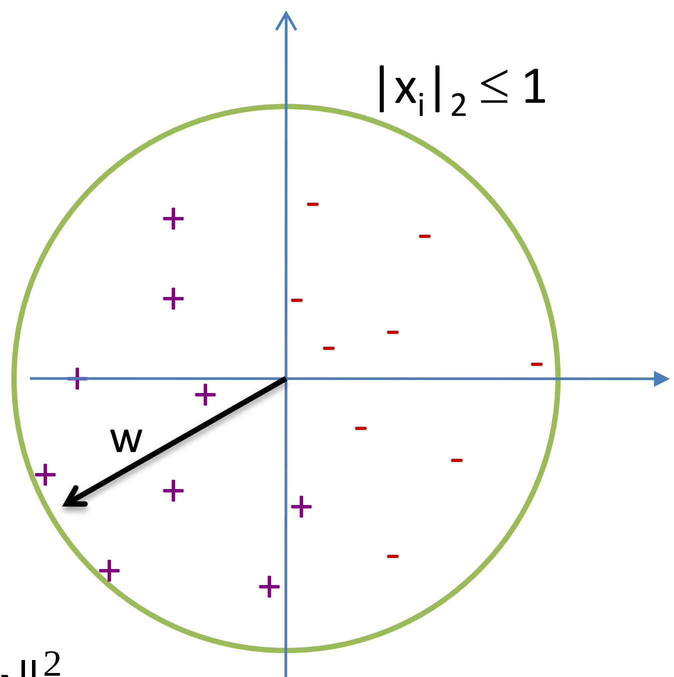
(attributes, label) = (x_i, y_i)

$x_i \in R^d$ y_i= 1 or -1 $i \in [n]$

For (x_i, y_i), w predicts:

w.p $P_i(1) = \dfrac{1}{1+e^{-<w,x_i>}}$ being positive

w.p $P_i(-1) = \dfrac{1}{1+e^{<w,x_i>}}$ being negative



Maximize the log-likelihood $\sum_i \log p_i(y_i)$

Minimize the log-loss $\sum_i \log \dfrac{1}{p_i(y_i)}$

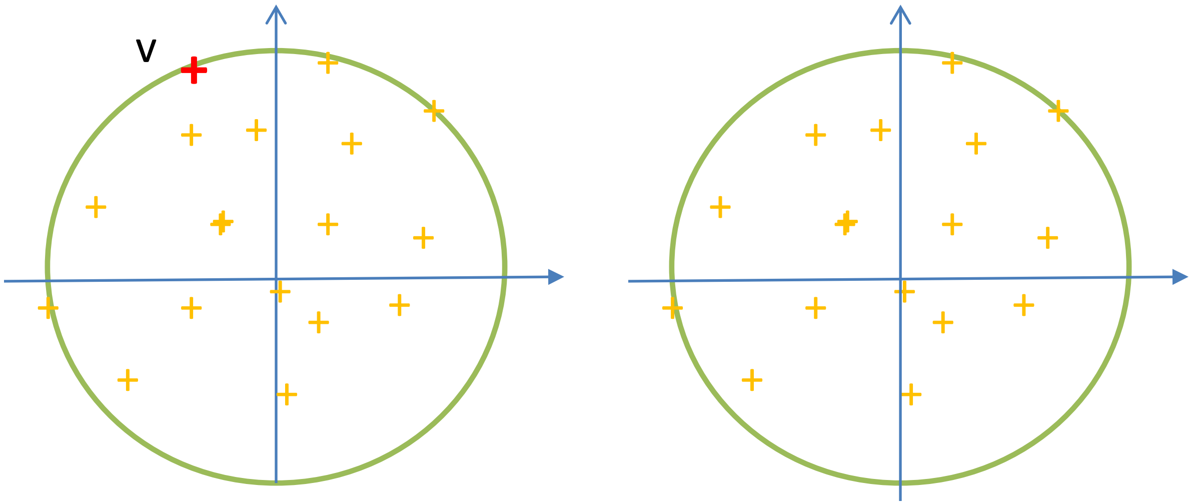
$$L^X(w) = \sum_i \log(1+e^{-y_i<w,x_i>}) \quad + \lambda \| w \|_2^2$$

Widely used in statistics, physics, social science, etc

Differentially Private

Logistic Regression

A LR algorithm $\mathbb{A}$ is ε-differentially private, if for all neighboring databases X and X',

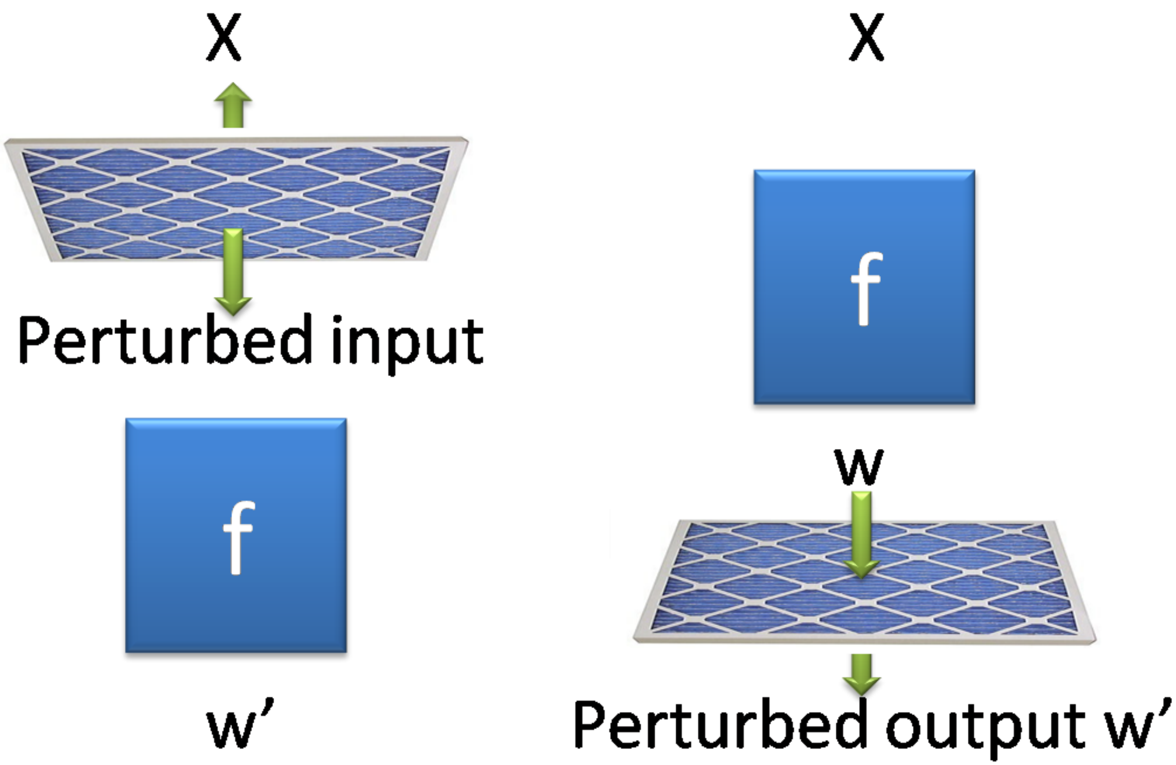


for all set of w, W

$$\Pr[\mathbb{A}(X) \in W] < e^{\varepsilon} \Pr[\mathbb{A}(X') \in W]$$

The behavior of the algorithm is "unchanged" no matter if a data point opts in or opts out.

Differentially Private LR

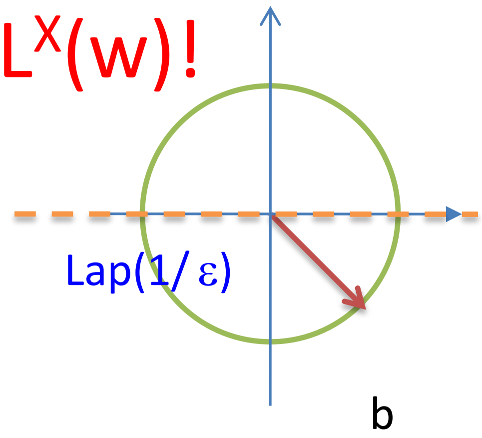


Gradient Descent [FM]   Add noise prop. to GS

$GS(LR) \leq 2/\lambda$ [CM]

Algorithm [CM]2

Perturb the Loss Function $L^X(w)$!



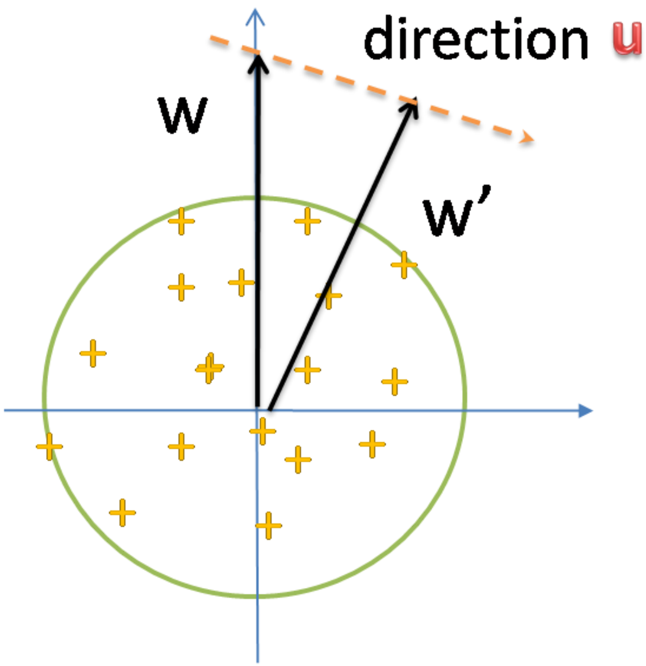
- Draw noise vector b w.p $p(b) \propto e^{-\varepsilon|b|}$
- Output w that minimizes

$$L^X(w) + \langle b, w \rangle$$

In [CM], simulation results show that [CM]2 outperforms the approach that adds noise proportional to GS.

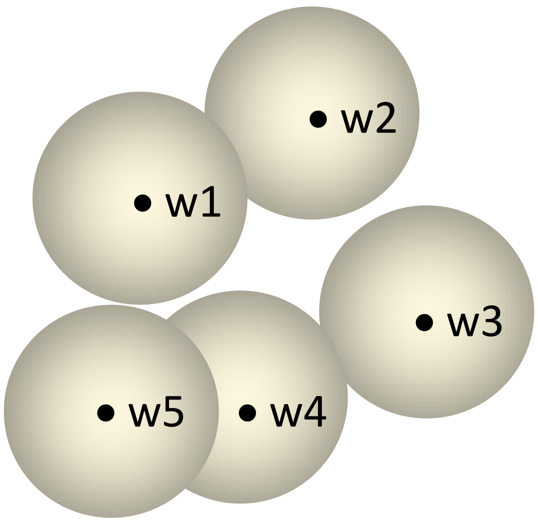
X: data set   w: true optimum   w': output of [CM]2

Noise = |w'-w | ???



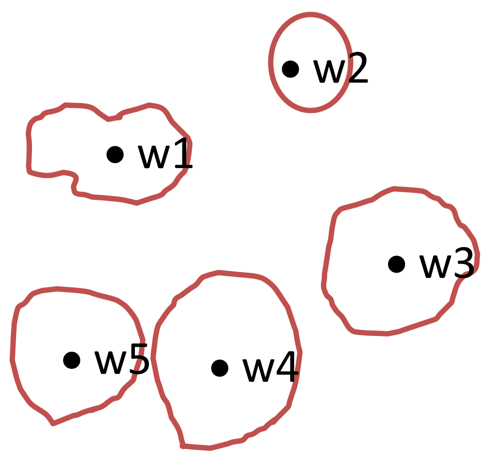
Sensitivity of Function

$$|Opt(X) - Opt(X')|$$

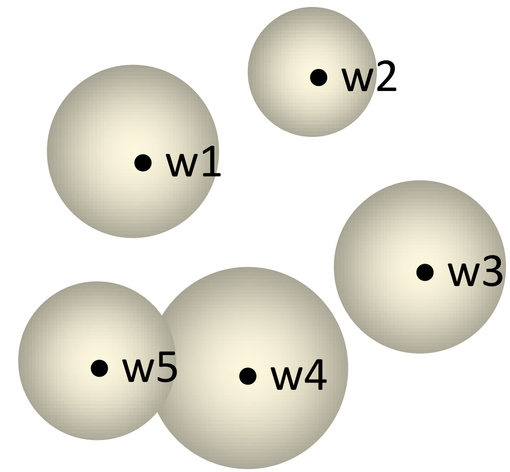


Global sensitivity

However,

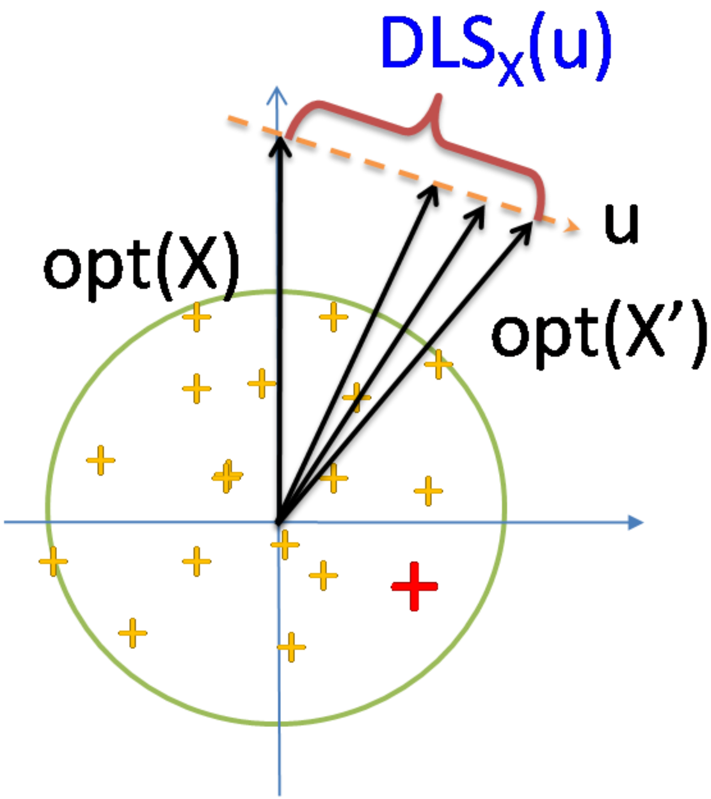


Real Sensitivity



local sensitivity

Directional Local Sensitivity



$DLS_X(u)$ is the supremum of $|opt(X') - opt(X)|$

over all neighboring data sets X', s.t. opt(X') -opt(X) is parallel to u.

$$DLS_X(u) \leq e/T_X(u)\ *$$

Summary

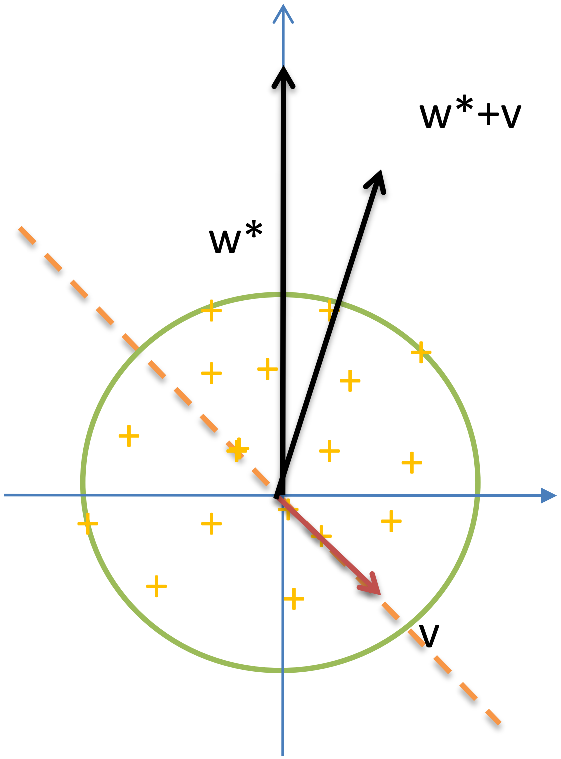
- Analysis of the CM2 algorithm.
- Directional local sensitivity.

Is the CM2 algorithm actually “tracing” DLS?

Algorithm [CM]1

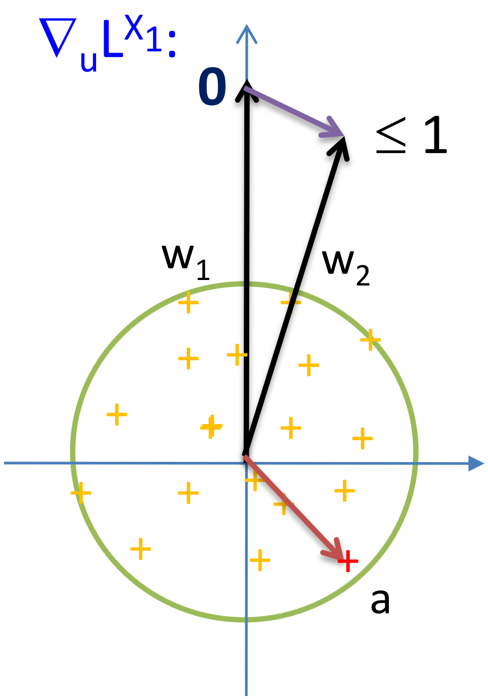
- Adding noise proportional to the global sensitivity.
  - Pick noise vector v with probability $p(v) \propto e^{-\varepsilon\lambda|v|}$
  - Output (w* + v)

On any line, the probability distribution of |v| is Lap(1/ $\varepsilon\lambda$)



Claim: GS(LR) $\leq 1/\lambda$

$X_2 = X_1 + (a, 1)$



$$L^{X_2}(w) = \underbrace{\lambda |w|^2 + \Sigma_i \log(1 + e^{-<x_i, w>})}_{L^{X_1}(w)} + \underbrace{\log(1 + e^{-<a, w>})}_{g^a(w)}$$

$\mathbf{0} = \nabla_u L^{X_2}(w_2) = \nabla_u L^{X_1}(w_2) \ldots$

$$\frac{<a,u>}{1+e^{<w_2,a>}} \leq 1$$

$\nabla_u L^{X_1}(w_2) = -\nabla_u g^a(w_2) \leq 1$

First derivative is the in...

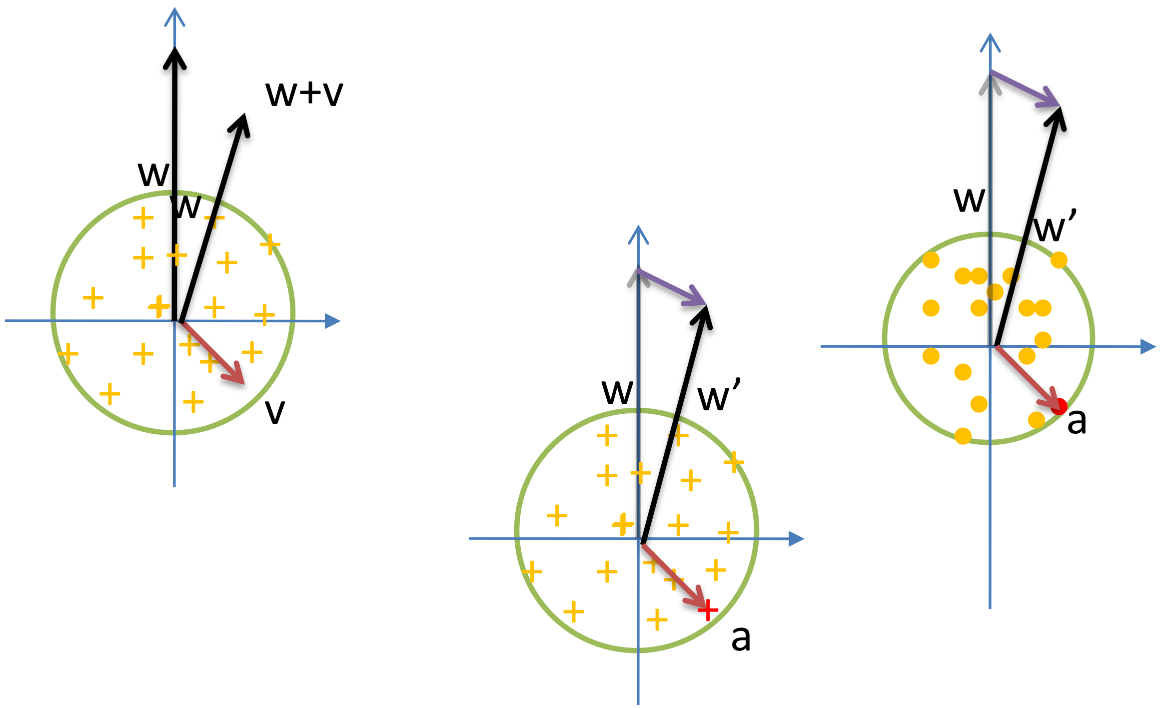
along the line

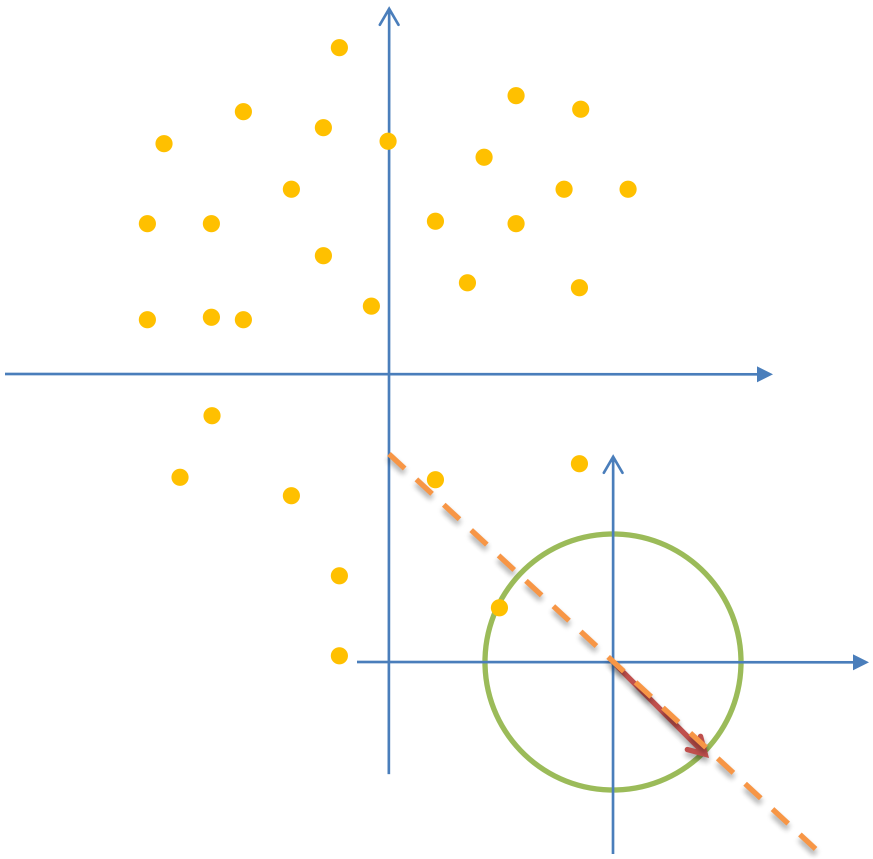
$$\sum_i \frac{<x_i,u>^2}{(1+e^{-<w,x_i>})(1+e^{<w,x_i>})} \geq 0$$

$(\nabla_u)^2 L^{X_1}(w) = \lambda + T \geq \lambda$

$1 \geq \nabla_u L^{X_1}(w_1) - \ldots \geq \lambda |w_1 - w_2|_2$

$|w_1 - w_2|_2 \leq 1/\lambda$


w
w+v
w
v
w
w’
a
w
w’
a



Differentially Private

Logistic Regression

Lots of previous works on differential privacy.

Add noise to achieve privacy

Tradeoff b/w privacy and utility

sensitivity of function

low sensitivity ➔ less noise

Roadmap

- Global sensitivity
  - [CM1]
- Local sensitivity
  - Our algorithm and [CM2]
- Directional local sensitivity
  - Add non-spherical noises
  - [CM2] and noisy gradient descent []