Roll: 001610501020

Q5(a) Optimization is the most essential ingredients component in machine learning algorithms. In most machine learning algorithms some kind of loss is defined and then this loss function is optimized by adjusting the weights. This loss function is optimized using some optimization routine. The choice optimization algorithm can make a difference between getting a good accuracy in hours or days.

Examples:

1. Logistic Regression:

Here the loss function is optimised.

Output = 0 or 1.

Hypothesis $Z = WX + B$

$h_\theta(x) = \text{sigmoid}(Z)$.

Now, we define the cost function.

$$\text{cost}(h_\theta(x), Y) = -\log(h_\theta(x)) \text{ if } y=1$$
$$= -\log(1-h_\theta(x)) \text{ if } y=0$$

This may be written as

$$C = \text{cost}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

∴ The formulation is

$$\underset{\theta}{\text{minimize}}\left(\text{cost}(h_\theta(x), y)\right)$$

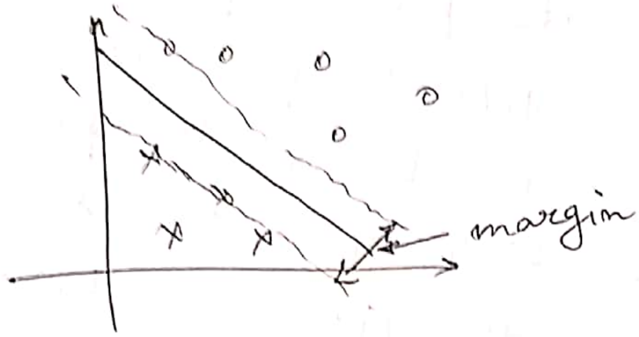
This may be done using gradient descent.

Anuran Chakraborty

Roll : 001610501020

2. SVM.

In SVM we optimize the ~~m~~ function so as to maximize the margin between the data points and the hyperplane. This loss function is known as hinge loss.



for $y_i = +1$, $\quad wx_i + b > 0$.

$y_i = -1$, $\quad wx_i + b < 0$.

Scaling we get

$y_i = 1 \quad wx_i + b > 1$ — (1)

$y_i = -1 \quad wx_i + b < -1$ — (2)

max. margin width is

$$M = (x^+ - x^-) \cdot n = (x^+ - x^-) \frac{w}{\|w\|} = \frac{2}{\|w\|}$$

$x^+ - x^- = 2$ from (1) and (2)

maximize $\frac{2}{\|w\|}$ such that

$wx_i + b > +1 \quad y_i = +1$

$wx_i + b < -1 \quad y_i = -1$

$$\boxed{\text{minimize } \frac{1}{2}\|w\|^2 \text{ such that } y_i(wx + b) \geq 1}$$

This can optimized using Quadratic Programming

Anuran Chakraborty

Roll : 001610501020

Q5 (b) when we solve a non-linear problem it is very important to know if the problem is convex or not. The convexity characteristics decide which solution method is suitable to use and what solution quality we can expect when applying this method. If the problem is convex then each local optimum will also be a global optimum. Since most methods are search methods that only guarantee to find a local optimum, the convexity properties decide if we with certainty or not can announce that the solution we have found is the global optimum.

if $f(x)$ is twice differentiable. then,

- $f(x)$ is an convex function if hessian matrix H is positive semi-definite for all $x \in X$.
- $f(x)$ is concave if H is $-$ve semi-definity $\forall x \in X$.

~~Say, for example $f(x)$ = $-x^2$~~

Anuran Chakraborty

$f(x_1, x_2) = -x_1^2 - 4x_2^2 + x_1x_2$ 001610501020

$\nabla = [-2x_1 + x_2 \quad -8x_2 + x_1]$

$$H = \begin{bmatrix} -2 & 1 \\ 1 & -8 \end{bmatrix}$$

Now find eigen value.

$|H - \lambda I| = 0$

or, $(-2-\lambda)(-8-\lambda) - 1 = 0.$

$\lambda^2 + 10\lambda + 15 = 0.$

$\therefore \lambda_1 = -5 + \sqrt{10} \quad \lambda_2 = -5 - \sqrt{10}.$

$\therefore$ All eigen values are < 0 $\therefore f(x_1, x_2)$ is not convex.

say $f(x_1, x_2) = x_1^2 + x_2^2$

$\nabla = [2x_1 \quad 2x_2]$

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$|H - \lambda I| = 0$

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = 0.$$

$\lambda = 2.$

$\therefore$ +ve definite H.

$\therefore f(x_1, x_2)$ is convex.

Anuran Chakraborty

Roll: 001610501020

Q5(c) In duality we ~~convert~~ consider each constraint as a variable and the RHS as its coefficient. Basically solving a dual problem is equivalent to solving its primal problem.

In LP models, the parameters are usually not exact. ~~In LP,~~ The parameters may change withing certain limits without changing the optimal solution. This is referred to as sensitivity analysis.

Relaxation is expanding the feasible region by making constraints "less restrictive".

- Removing a constraint
- Increasing RHS of $\leq$ constraint
- Decreasing RHS of $\geq$ constraint.

When we relax a binding constraint, the optimal objective function value will improve or stay the same. For non-~~bin~~ binding it remains same.

Restriction is contracting the feasible region by making constraints more restrictive.

- Adding a new constraint.
- Decreasing RHS $\leq$ constraint
- Increasing RHS $\geq$ constraint.

Objective functions worsens or remains same

Anuran Chakraborty

Scanned with CamScanner