(1)

$$f(y) - f(x) = g(y) - g(x) + \sum \left(h_i(y_i) - h_i(x_i) \right)$$

$$\geq \nabla g(x)^T (y - x) + \sum \left(h_i(y_i) - h_i(x_i) \right)$$

$$= \sum_{i=1}^{n} \underbrace{\left(\nabla_i g(x) \cdot (y_i - x_i) + h_i(y_i) - h_i(x_i) \right)}_{\geq 0}$$

$$\geq 0$$

$$F_i(x_i) = g(x_i; x_{-i}) + h_i(x_i)$$

$$0 \in \partial F_i(x_i)$$

$$0 \in \{\nabla_i g(x)\} + \partial h_i(x_i)$$

$$\nabla_i g(x) \in -\partial h_i(x_i)$$

thus $h_i(y_i) \geq h_i(x_i) - \nabla_i g(x) \cdot (y_i - x_i)$

by def'n.

proves the claim.

$$X^T X \beta = X^T y$$

$$\frac{X_i^T (y - X_{-i} \beta_{-i})}{X_i^T X_i} = \frac{X_i^T (y - X\beta)}{X_i^T X_i} + \frac{X_i^T \cancel{X_i} \beta_i}{X_i^T \cancel{X_i}}$$

$$= \frac{X_i^T r}{X_i^T X_i} + \beta_i$$

(2)

$$\min_{\beta} \; \underset{\uparrow}{\ell(X\beta)} + \lambda \|\beta\|_1$$

logistic regression loss

Note: $\nabla \ell(X\beta) = X^T(y - p(\beta))$

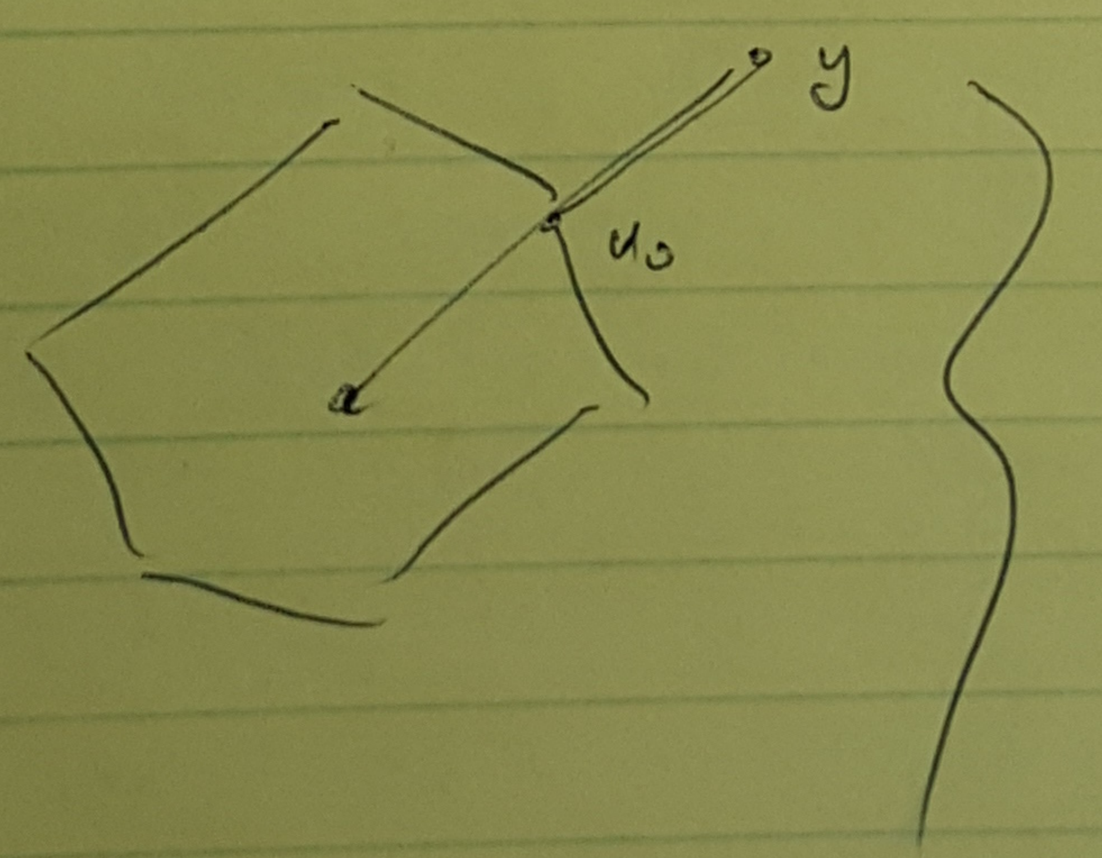
where $p_i(\beta) = \dfrac{1}{1+\exp(-(X\beta)_i)}$ $\quad i = 1, \dots n$

Therefore KKT conditions

$$X_1^T(y - p(X\beta)) = \lambda s_1$$
$$X_2^T(y - p(X\beta)) = \lambda s_2$$
$$\vdots$$
$$X_p^T(y - p(X\beta)) = \lambda s_p.$$

$A = \{ j : \; KKT_j \text{ does not hold} \}$

$$\min_{\beta} \frac{1}{2} \sum \left(y_i - \sum_{j=1}^{p} X_{ij} \beta_j\right)^2 + \lambda \|\beta\|_1$$



$$\min_{u} \; g(u)$$
$$|X_i^T u| \leq 1 \quad i = 1 \dots p$$
$$g(u) \geq \gamma$$