Formula

Week 1

- **Logit(aka. log-odds)**: the logarithm of the odds:

  $$Logit(p_1) = \vec{w}^T \vec{x}$$

  where $logit(p_1) = \ln(\frac{p_1}{1-P_1})$

  > Logit enables us to map from **[0,1]** to **[$-\infty, \infty$]**
  > If $logit(p_1) \geq 0$, predict `class 1`
  > If $logit(p_1) < 0$, predict `class 0`

- $$p_1 = \frac{e^{\vec{w}^T \vec{x}}}{1 + e^{\vec{w}^T \vec{x}}}$$

  $$p_0 = 1 - p_1 = \frac{1}{1 + e^{\vec{w}^T \vec{x}}}$$

- **Likelihood function**

  $$\begin{aligned} \prod_{i=1}^{N} P_{y^i} &= \prod_{i=1}^{N} p(y^i | x^i; \vec{w}) = p(\vec{y} | \vec{X}, \vec{w}) = L(\vec{w}) && (1) \\ &= \underbrace{\prod_{i=1}^{N} p(1|\vec{x}^i, \vec{w})^{(y_i)} (1 - p(1|\vec{x}^i, \vec{w}))^{(1-y_i)}}_{\text{这一段使用了Bernoulli distribution进行转换}} && (2) \end{aligned}$$

- **Log-Likelihood**

  $$\ln(L(\vec{w})) = \ln \prod_{i=1}^{N} P_{y^i} = \sum_{i=1}^{N} \ln P_{y^i}$$

- **Loss Function**

  $$\begin{aligned} E(\vec{w}) &= -\ln(L(\vec{w})) = -\sum_{i=1}^{N} \ln P_{y^i} && (3) \\ &= -\sum_{i=1}^{N} y^i \ln p(1|\vec{x}^i, \vec{w}) + (1 - y^i) \ln(1 - p(1|\vec{x}^i, \vec{w})) && (4) \end{aligned}$$

- Gradient descent adjusts $\vec{w}$ iteratively in the direction that leads to the biggest decrease (steepest descent) in $E(\vec{w})$.

$$x := x - \eta \frac{df}{dx}$$

即

$$\vec{w} = \vec{w} - \eta \nabla E(\vec{w})$$

where $\eta > 0$ and $\nabla E(\vec{w}) = \sum_{i=1}^{N} (p(1|\vec{x}^i, \vec{w}) - y^i)\vec{x}^i$

Week 2

- Newton-Raphson

$$w = w - \frac{E'(w)}{E''(w)}$$

- `Taylor Polynomial of degree` n can be used to approximate a function $E(w)$ at w_0:

$$T_n(w) = \sum_{k=0}^{n} \frac{E^k(w_0)}{k!}(w - w_0)^k$$

  where $E^{(k)}(w_0)$ is the k-th order derivative of E at w_0

- Weight Update Rule
  - Univariate update rule:

$$w = w - \frac{E'(w)}{E''(w)}$$

  - Multivariate update rule:

$$\vec{w} = \vec{w} - H_E^{-1}(\vec{w}) \nabla E(\vec{w})$$

    where $H_E^{-1}(\vec{w})$ is the inverse of the Hessian at the old $\vec{w}$ and $\nabla E(\vec{w})$ is the gradient at the old w

- Logistic Regression - Iterative Reweighted Least Squares

$$\vec{w} = \vec{w} - H_E^{-1}(\vec{w}) \nabla E(\vec{w})$$

$$H_E(\vec{w}) = \sum_{i=1}^{N} p(1|\vec{x}^{(i)}, \vec{w})(1 - p(1|\vec{x}^{(i)}, \vec{w}))\vec{x}^{(i)}\vec{x}^{(i)^T}$$

$$\nabla_E(\vec{w}) = \sum_{i=1}^{N} (p(1|\vec{x}^{(i)}, \vec{w}) - y^{(i)})\vec{x}^{(i)}$$

- Adopting Nonlinear Transformations in Logistic Regression

$$logit(p_1) = \vec{w}^T \phi(\vec{x})$$

$$p_1 = p(1|\phi(\vec{x}), \vec{w}) = \frac{e^{\vec{w}^T\phi(\vec{x})}}{1 + e^{\vec{w}^T\phi(\vec{x})}}$$

$$\text{Given } \; J = \{(\phi(\vec{x}^{(1)}, y^1), (\phi(\vec{x}^{(2)}, y^2), ..., (\phi(\vec{x}^{(N)}, y^N)\}, \arg\min_{\vec{w}} E(\vec{w})$$

$$E(\vec{w}) = -\sum_{i=1}^{N} y^i \ln p(1|\phi(\vec{x}^{(i)}, \vec{w}) + (1 - y^i)\ln(1 - p(1|\phi(\vec{x}^{(i)}, \vec{w}))$$

$$\triangledown E(\vec{w}) = \sum_{i=1}^{N} (p(1|\phi(\vec{x}^{(i)}), \vec{w}) - y^i)\phi(\vec{x}^{(i)}$$

$$H_E(\vec{w}) = \sum_{i=1}^{N} p(1|\phi(\vec{x})^i, \vec{w})(1 - p(1|\phi(\vec{x}^{(i)}), \vec{w}))\phi(\vec{x}^{(i)})\phi(\vec{x}^{(i)})^T$$

Week 3

- Perpendicular Distance From a Point $\vec{x}^{(n)}$ to a Hyperplane $h(\vec{x}) = 0$

$$dist(h, \vec{x}^{(n)}) = \frac{|h(\vec{x}^{(n)})|}{||\vec{w}||} = \frac{y^{(n)}h(\vec{x}^{(n)})}{||\vec{w}||}$$

where $||w|| = \sqrt{\vec{w}^T\vec{w}}$ is the Euclidean norm(the length of the vector $\vec{w}$)

> $$\min_n dist(h, \vec{x}^{(n)})$$
> $$\downarrow$$
> $$\arg\max_{\vec{w},b}\{\min_n dist(h, \vec{x}^{(n)})\}$$
>
> `Constraint:`
>
> $$Subject\ to\ y^{(n)}h(\vec{x}^{(n)}) > 0, \forall(\vec{x}^{(n)}, y^{(n)}) \in J$$

> $$\arg\max_{\vec{w},b}\{\min_n(\frac{y^{(n)}h(\vec{x}^{(n)})}{||\vec{w}||})\}$$
> $$\downarrow$$
> $$\arg\max_{\vec{w},b}\{\frac{1}{||\vec{w}||}\min_n(y^{(n)}h(\vec{x}^{(n)}))\}$$
>
> `Constraint:`
>
> $$Subject\ to\ y^{(n)}h(\vec{x}^{(n)}) > 0, \forall(\vec{x}^{(n)}, y^{(n)}) \in J$$
> $$Subject\ to\ \min_n y^{(n)}h(\vec{x}^{(n)}) = 1, \forall(\vec{x}^{(n)}, y^{(n)}) \in J$$

> $$\arg\max_{\vec{w},b}\{\frac{1}{||\vec{w}||}\}$$

$$\downarrow$$

$$\arg\min_{\vec{w},b}\{||\vec{w}||\}$$

Constraint:

$$Subject\ to\ \min_{n} y^{(n)}h(\vec{x}^{(n)}) = 1, \forall(\vec{x}^{(n)}, y^{(n)}) \in J\ stricter$$
$$Subject\ to\ y^{(n)}h(\vec{x}^{(n)}) \geq 1, \forall(\vec{x}^{(n)}, y^{(n)}) \in J\ looser$$

$$\arg\min_{\vec{w},b}\{\frac{1}{2}||\vec{w}||^2\}$$

Constraint:

$$Subject\ to\ y^{(n)}(\vec{w}^T\phi(\vec{x}^{(n)} + b) \geq 1, \forall(\vec{x}^{(n)}, y^{(n)}) \in J$$