| | Model (h) | Loss [J(w)] | Opt-Normal equation (w) | Opt-Weight update |
|---|---|---|---|---|
| Linear Reg | $\mathbf{w}^T\mathbf{x}$ | $\frac{1}{2}\left(\mathbf{Xw}-\mathbf{y}\right)^T\left(\mathbf{Xw}-\mathbf{y}\right)$ | $\mathbf{X}^{-1}\mathbf{y}$ | $\mathbf{w}_{k+1} := \mathbf{w}_k - \alpha\mathbf{X}^T\left(\mathbf{Xw}-\mathbf{Y}\right)$ |
| Poly Reg | $\mathbf{w}^T\phi(\mathbf{x})$ | $\frac{1}{2}\left(\mathbf{Xw}-\mathbf{y}\right)^T\left(\mathbf{Xw}-\mathbf{y}\right)$ | $\mathbf{X}^{-1}\mathbf{y}$ | $\mathbf{w}_{k+1} := \mathbf{w}_k - \alpha\mathbf{X}^T\left(\mathbf{Xw}-\mathbf{Y}\right)$ |
| Ridge | | $\frac{1}{2}(\mathbf{Xw}-\mathbf{y})^T(\mathbf{Xw}-\mathbf{y})+\frac{\lambda}{2}\mathbf{w}^T\mathbf{w}$ | $\left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}$ | $\mathbf{w}_{k+1} := \mathbf{w}_k - \alpha\left\{\mathbf{X}^T(\mathbf{Xw}-\mathbf{y})+\lambda\mathbf{w}\right\}$ |
| Lasso | | $\frac{1}{2}(\mathbf{Xw}-\mathbf{y})^T(\mathbf{Xw}-\mathbf{y})+\frac{\lambda}{2}\sum_{j=1}^{m}\lvert w_j\rvert$ | Use sklearn | Use sklearn |
| Least squares | | $\frac{1}{2}\left(\mathbf{Xw}-\mathbf{y}\right)^T\left(\mathbf{Xw}-\mathbf{y}\right)$ | $\left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}$ | $\mathbf{w}_{k+1} := \mathbf{w}_k - \alpha\left\{\mathbf{X}^T(\mathbf{Xw}-\mathbf{y})+\lambda\mathbf{w}\right\}$ |
| Perceptron | $\mathrm{sign}(\mathbf{w}^T\phi(\mathbf{x}))$ | $\sum_{i=1}^{n}\max(0,-y^{(i)}h_{\mathbf{w}}(x^{(i)}))$. | $\left(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}$ | $\mathbf{w}^{(t+1)} := \mathbf{w}^{(t)}+\alpha\left(y^{(i)}-\hat{y}^{(i)}\right)\phi(\mathbf{x}^{(i)})$<br>[For one example, sum up for all such examples] |
| Logistic Reg | $\frac{1}{1+\exp(-\mathbf{w}^T\mathbf{x})}$ | $-\sum_{i=1}^{n}y^{(i)}\log\left(h(\mathbf{x}^{(i)})\right)+(1-y^{(i)})\log\left(1-h(\mathbf{x}^{(i)})\right)$ | | $\mathbf{w} := \mathbf{w}-\alpha\left(\mathbf{X}^T\left(g(\mathbf{Xw})-\mathbf{y}\right)\right)$ |
| Softmax | $\begin{bmatrix}\frac{\exp(\mathbf{w}_1^T\mathbf{x})}{\sum_{j=1}^{k}\exp(\mathbf{w}_j^T\mathbf{x})}\\ \frac{\exp(\mathbf{w}_2^T\mathbf{x})}{\sum_{j=1}^{k}\exp(\mathbf{w}_j^T\mathbf{x})}\\ \vdots\\ \frac{\exp(\mathbf{w}_{k-1}^T\mathbf{x})}{\sum_{j=1}^{k}\exp(\mathbf{w}_j^T\mathbf{x})}\end{bmatrix}$ | | | $\mathbf{w} := \mathbf{w}-\alpha\left(\mathbf{X}^T\left(h_{\mathbf{w}}(\mathbf{X})-\mathbf{y}\right)\right)$<br>NOTE: Class label of the highest value of softmax is assigned to the sample (will replace h in the above equation) |
| Naive Bayes | | $\sum_{i=1}^{n}\log p(y^{(i)};\mathbf{w})+\sum_{i=1}^{n}\sum_{j=1}^{m}\log p(x_j^{(i)}\lvert y^{(i)};\mathbf{w})$<br>NOTE: ***prior*** is given as $P(y=c)=\frac{\sum_{i=1}^{n}1(y^{(i)}=c)}{n}$<br>and $p(x_j\lvert y_c)$ depends on the distribution | | |
| SVM | $\mathrm{sign}(\mathbf{w}^T\mathbf{x}+b)$ | $\sum_{i=1}^{n}\alpha^{(i)}-\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}\langle\alpha^{(i)}y^{(i)}x^{(i)},\alpha^{(k)}y^{(k)}x^{(k)}\rangle$ | | $\mathbf{w}^{(\text{new})}=\mathbf{w}^{(\text{old})}-\text{learning rate}\times\left(\mathbf{w}+C\sum_{i=0}^{n}\mathbf{1}\left(1-\mathbf{y}^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}+b)>0\right)\mathbf{y}^{(i)}\mathbf{x}^{(i)}\right)$<br>$b^{(\text{new})}=b^{(\text{old})}-\text{learning rate}\times C\sum_{i=0}^{n}\mathbf{1}\left(1-\mathbf{y}^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}+b)>0\right)\mathbf{y}^{(i)}$ |
| ID3 | $p_{i,k}=\frac{1}{N_i}\sum_{x^{(i)}\in R_i}1(y^{(i)}=k)$ | $\mathrm{Entropy}(S)-\sum_{v\in Values(A)}\frac{\lvert S_v\rvert}{\lvert S\rvert}\mathrm{Entropy}(S_v)$<br>where Entropy $H_i=-\sum_{k=1}^{n}p_{i,k}\log_2 p_{i,k}$ | | |
| CART | $\hat{y}=\sum_{i=1}^{M}c_i 1(x\in R_i)$ | $\min_{j,s}\left[\min_{c_1}\sum_{i;\mathbf{x}^{(i)}\in R_1}(\mathbf{y}^{(i)}-c_1)^2+\min_{c_2}\sum_{i;\mathbf{x}^{(i)}\in R_2}(\mathbf{y}^{(i)}-c_2)^2\right]$ | | |
| GradientBoost | $\boxed{y_{train_{p0}}=\frac{1}{n}\sum_{i=1}^{n}y_{train_i}}$ | $\boxed{r_0=y_{train}-y_{train_{p0}}}$ | | $\boxed{y_{train_{p1}}=y_{train_{p0}}+\alpha f_0(X_{train})}$ |
| AdaBoost | | | | $\text{weight}^{(\text{new})}=\text{weight}^{(\text{old})}\times e^{\alpha}$ (incorrectly classified)<br>$\text{weight}^{(\text{new})}=\text{weight}^{(\text{old})}\times e^{-\alpha}$ (correctly classified)<br>NOTE: $\alpha=\frac{1}{2}\ln\frac{1-\text{Total Error}}{\text{Total Error}}$ |
| K-Means | $\mu^{(r)}=\frac{1}{\lvert\mathbf{x}^{(i)}\in c_r\rvert}\sum_{i=1}^{n}\mathbf{1}(\mathbf{x}^{(i)}\in c_r)\mathbf{x}^{(i)}$ | $J(c)=\sum_{r=1}^{k}\sum_{i=1}^{n}\mathbf{1}(\mathbf{x}^{(i)}\in c_r)(\lvert\lvert\mathbf{x}^{(i)}-\mu^{(r)}\rvert\rvert)^2$ | | |