Prediction of MHC Class I and II binding peptides incorporating bayesian transfer hierarchies

Ravikiran Janardhana

December 4, 2012

So far...

- Increased feature set for Elastic Net (baseline) by incorporating interaction features, i.e, Protein (a,b) at Pos (x,y). There are 80100 features now compared to 300 earlier
- Using Matlab Elastic Net implementation (Lasso) and SVM, the accuracies are now comparable, varies from 70-75%, the state of the art reports 80% for real data. The accuracy of Elastic Net earlier was mediocre at 53% and there is a drastic improvement with the addition of these features.
- **Question:** How many training samples and testing samples needs to be there in each set for acceptable result?

Model:

The optimization problem for two related MHC-Class II alleles classifier is given by

$$\begin{aligned}\underset{\mathbf{w}^1,\mathbf{w}^2,\mathbf{w}}{\text{minimize}} \quad & \frac{1}{2}\left\|\mathbf{y}_1-\mathbf{x}_1\right\|_2^2+\frac{1}{2}\left\|\mathbf{y}_2-\mathbf{x}_2\right\|_2^2+ \\ & \lambda_1\left\|\mathbf{w}^1\right\|_1+\lambda_2\left\|\mathbf{w}^2\right\|_1+\alpha\left\|\mathbf{D}\mathbf{w}\right\|_1.\end{aligned}$$

where,

$$\mathbf{D}=\begin{bmatrix}1000\,\ldots\ldots\ldots\ -1000\\ 1000\,\ldots\ldots\ldots\ 0-100\\ 1000\,\ldots\ldots\ldots\ 00-10\\ 1000\,\ldots\ldots\ldots\ 000-1\end{bmatrix} \qquad \mathbf{w}=\begin{bmatrix}\mathbf{w}^1\\ \mathbf{w}^2\end{bmatrix}.$$

We are going to introduce new variables $\mathbf{z}^1,\mathbf{z}^2,\mathbf{z}^3,\mathbf{z}^4,\mathbf{z}^5$ and reformulate the problem

$$\begin{aligned}\underset{\mathbf{w},\mathbf{z}^1,\mathbf{z}^2,\mathbf{z}^3,\mathbf{z}^4,\mathbf{z}^5}{\text{minimize}} \quad & \frac{1}{2}\left\|\mathbf{y}_1-\mathbf{z}_1\right\|_2^2+\frac{1}{2}\left\|\mathbf{y}_2-\mathbf{z}_2\right\|_2^2+ \\ & \lambda_1\left\|\mathbf{z}^3\right\|_1+\lambda_2\left\|\mathbf{z}^4\right\|_1+\alpha\left\|\mathbf{z}^5\right\|_1.\end{aligned}$$

Writing out the augmented lagrangian for the above problem,

$$\begin{aligned}\mathrm{AL}(\mathbf{w},\mathbf{z}^0,\mathbf{z}^1,\mathbf{z}^2,\mathbf{z}^3,\mathbf{z}^4,\mathbf{z}^5,\mathbf{u}^1,\mathbf{u}^2,\mathbf{u}^3,\mathbf{u}^4,\mathbf{u}^5) \;=\; & \frac{1}{2}\left\|\mathbf{y}_1-\mathbf{z}_1\right\|_2^2+\frac{1}{2}\left\|\mathbf{y}_2-\mathbf{z}_2\right\|_2^2+\lambda_1\left\|\mathbf{z}^3\right\|_1+\lambda_2\left\|\mathbf{z}^4\right\|_1+\alpha\left\|\mathbf{z}^5\right\|_1 \\ & +\mathbf{u}^1(\mathbf{z}^1-\mathbf{x}^1)+\mathbf{u}^2(\mathbf{z}^2-\mathbf{x}^2) \\ & +\mathbf{u}^3(\mathbf{z}^3-\mathbf{w}^1)+\mathbf{u}^4(\mathbf{z}^4-\mathbf{w}^2)+\mathbf{u}^5(\mathbf{z}^5-\mathbf{D}\mathbf{w}) \\ & +\frac{\rho}{2}\left\|\mathbf{z}^1-\mathbf{x}^1\right\|_2^2+\frac{\rho}{2}\left\|\mathbf{z}^2-\mathbf{x}^2\right\|_2^2 \\ & +\frac{\rho}{2}\left\|\mathbf{z}^3-\mathbf{w}^1\right\|_2^2+\frac{\rho}{2}\left\|\mathbf{z}^4-\mathbf{w}^2\right\|_2^2+\frac{\rho}{2}\left\|\mathbf{z}^5-\mathbf{D}\mathbf{w}\right\|_2^2\end{aligned}$$