A few slides on Interior Point SVM entries

For PASCAL Largescale Learning Challenge workshop

Kristian Woodsend

School of Mathematics
University of Edinburgh

4th July 2008

Linear SVM using interior point method

Primal formulation

$$\min_{\underline{w}, w_0, \underline{\xi}} \quad \frac{1}{2}\underline{w}^T\underline{w} + C\underline{e}^T\underline{\xi}$$

$$\text{s.t.} \quad Y(X^T\underline{w} + w_0\underline{e}) \geq \underline{e} - \underline{\xi}$$

$$\underline{\xi} \geq \underline{0}$$

Very many constraints $\Rightarrow \mathcal{O}(n^2)$

Dual formulation

$$\min_{\underline{\alpha}} \quad \frac{1}{2}\underline{\alpha}^T \, YX^TXY \, \underline{\alpha} - \underline{e}^T\underline{\alpha}$$

$$\text{s.t.} \quad \underline{y}^T\underline{\alpha} = 0$$

$$\underline{0} \leq \underline{\alpha} \leq C\underline{e}$$

Dense Hessian $\Rightarrow \mathcal{O}(n^3)$

Linear SVM using interior point method

Primal formulation

$$\min_{\underline{w}, w_0, \underline{\xi}} \quad \frac{1}{2}\underline{w}^T\underline{w} + C\underline{e}^T\underline{\xi}$$
$$\text{s.t.} \quad Y(X^T\underline{w} + w_0\underline{e}) \geq \underline{e} - \underline{\xi}$$
$$\underline{\xi} \geq \underline{0}$$

Very many constraints $\Rightarrow \mathcal{O}(n^2)$

Dual formulation

$$\min_{\underline{\alpha}} \quad \frac{1}{2}\underline{\alpha}^T \, YX^TXY \, \underline{\alpha} - \underline{e}^T\underline{\alpha}$$
$$\text{s.t.} \quad \underline{y}^T\underline{\alpha} = 0$$
$$\underline{0} \leq \underline{\alpha} \leq C\underline{e}$$

Dense Hessian $\Rightarrow \mathcal{O}(n^3)$

Separable formulation

$$\min_{\underline{w}, \underline{\alpha}} \quad \frac{1}{2}\underline{w}^T\underline{w} - \underline{e}^T\underline{\alpha}$$
$$\text{s.t.} \quad \underline{w} - XY\underline{\alpha} = \underline{0}$$
$$\underline{y}^T\underline{\alpha} = 0$$
$$\underline{0} \leq \underline{\alpha} \leq C\underline{e}.$$

Based on dual formulation.

Few constraints $\Rightarrow \mathcal{O}(m^2)$

Diagonal Hessian $\Rightarrow \mathcal{O}(n)$

Direct access to $\underline{w}$ and w_0 using primal-dual IPM

Efficient computations on multi-core processors

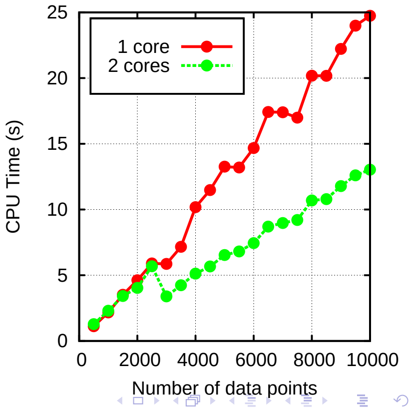
Vast majority of time ($>$ %95) is spent calculating matrix M:

$$M \equiv \begin{bmatrix} I_m & -XY \\ 0 & \underline{y}^T \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & \Theta_\alpha^{-1} \end{bmatrix}^{-1} \begin{bmatrix} I_m & 0 \\ -YX^T & \underline{y} \end{bmatrix}$$

Θ_α is diagonal matrix associated with logarithmic barrier.

Aim to minimize number of layer 2 cache misses:

- Divide data into “cache-sized” blocks
- Use BLAS Level 3 operations as much as possible (here DGEMM)
- BLAS libraries support shared memory threading across processor cores e.g. GotoBLAS, Intel MKL
- Layer 2 cache misses $\approx$ 2% of data reads



Termination based on hyperplane stability

- $\underline{w}$ (hyperplane) converges remarkably quickly
- $\underline{\alpha}$, duality gap and residual errors converge afterwards
- Use hyperplane stability for termination in version 2

Angle ϕ of hyperplane between iteration $i-1$ and i

$$\cos\phi = \frac{(\underline{w}^{(i-1)})^T \underline{w}^{(i)}}{\|\underline{w}^{(i-1)}\| \|\underline{w}^{(i)}\|}$$

Precision recall and change in hyperplane angle for alpha dataset: