О выборе наилучшего квадратичного регуляризатора в обобщенных линейных моделях классификации

Ветров Д. П., Кропотов Д. А.
VetrovD@yandex.ru, DKropotov@yandex.ru
Москва, ВМиК МГУ, ВЦ РАН

Обобщенные линейные модели (generalized linear models) в последние годы являются популярным средством решения задач классификации и восстановления регрессии. Примерами таких моделей могут служить методы опорных и релевантных векторов, логистическая регрессия, и др. Настройка весов производится путем оптимизации суммы некоторого функционала качества, связанного с ошибкой на обучающей выборке, и регуляризатора, предотвращающего перенастройку на данные.

Рассмотрим стандартную задачу классификации на два класса по заданной обучающей выборке $(X, T) = \{\mathbf{x}_i, t_i\}_{i=1}^m$, где $\mathbf{x} \in \mathbb{R}^d$, а $t \in \{-1, 1\}$.

Статистические модели обучения

При статистическом подходе в качестве функционала, связанного с ошибкой на обучении, используется логарифм правдоподобия правильной классификации обучающей выборки

$$L(T|X, \mathbf{w}) = -\sum_{i=1}^{m} \log\left(1 + \exp\left(-t_i \sum_{j=1}^{N} w_j \varphi_j(\mathbf{x}_i)\right)\right), \tag{1}$$

где $\{\varphi_j(\mathbf{x})\}_{j=1}^N$ — базисные функции, зафиксированные до начала обучения. Настройка весов производится путем оптимизации суммы (1) и регуляризатора, штрафующего большие значения весов $\mathbf{w}$ во избежание перенастройки на данные. Наиболее популярным регуляризатором является квадратичный, в простейшем случае имеющий вид

$$R_l(\mathbf{w}, \lambda) = -\lambda \mathbf{w}^{\mathrm{T}} I \mathbf{w} = -\lambda \sum_{j=1}^{N} w_j^2.$$

По такой схеме работает классический метод логистической регрессии, в котором коэффициент λ подбирается с помощью трудоемкой процедуры кросс-валидации. Популярность квадратичного регуляризатора обуславливается, помимо прочего, тем, что он соответствует ридж-регуляризации гессиана (часто вырожденного или плохо обусловленного) логарифмического правдоподобия, облегчающей процедуру оптимизации весов.

В методе релевантных векторов [3] используется более сложный регуляризатор вида

$$R_r(\mathbf{w}, \Lambda) = -\mathbf{w}^{\mathrm{T}}\Lambda\mathbf{w} = -\sum_{j=1}^{N} \lambda_j w_j^2,$$

где Λ — неотрицательная диагональная матрица. Таким образом, каждому весу присваивается собственный коэффициент регуляризации. Выбор значений λ_j проводится с помощью процедуры байесовского обучения путем максимизации *обоснованности* модели

$$\Lambda = \arg\max E(\Lambda) = \arg\max \int \exp\left(L(T|X, \mathbf{w}) + R_r(\mathbf{w}, \Lambda)\right) d\mathbf{w}.$$

Недиагональная регуляризация

Авторы предлагают рассмотреть более общий случай произвольной неотрицательно определенной матрицы регуляризации

$$R_i(\mathbf{w}) = -\mathbf{w}^{\mathrm{T}}A\mathbf{w}, \quad A^{\mathrm{T}} = A, \quad A \geqslant 0.$$

Настройка матрицы A производится также путем оптимизации обоснованности модели $E(A)$.

Пусть $\mathbf{w}_{ML} = \arg\max L(T|X, \mathbf{w})$, $H = -\nabla\nabla L(T|X, \mathbf{w})|_{\mathbf{w}=\mathbf{w}_{ML}} \geqslant 0$. Обозначим $M = H(H + A)^{-1}A$. Тогда можно показать [1], что

$$\frac{\partial \log E(A)}{\partial A} = -0.5\left[M^{-1} - \mathbf{w}_{ML}\mathbf{w}_{ML}^{\mathrm{T}}\right].$$

Приравнивая производную нулю отсюда можно получить

$$A^{-1} = \mathbf{w}_{ML}\mathbf{w}_{ML}^{\mathrm{T}} - H^{-1}.$$

Матрица A является симметричной, но имеет не более одного положительного собственного значения. В силу симметрии существует ортогональная матрица U, такая что $D = U^{\mathrm{T}}A^{-1}U$ является диагональной матрицей. Заменяя все отрицательные собственные значения матрицы D^{-1} на $+\infty$ (что соответствует наибольшему значению обоснованности среди неотрицательных собственных значений в случае, когда экстремум достигается в отрицательной области) получаем, что матрица регуляризации A будет штрафовать с бесконечно большим коэффициентом все значения весов $\mathbf{w}$, кроме тех, которые лежат вдоль вектора $\mathbf{u}$, соответствующего единственному положительному собственному значению d^{-1}

Задача	**SVM**	**RVM**	**LogReg**	**IREVM**	**VRVM**
Bupa	28.46	33.33	57.97	29.68	30.78
Heart	19.33	18.15	22.44	18.00	17.56
Hepatitis	17.55	14.32	19.48	16.26	13.03
Votes	4.78	5.56	5.98	4.92	5.61
WPBC	21.72	23.64	23.84	21.11	24.04
Laryngeal1	17.75	16.81	19.44	17.28	17.37
Weaning	10.99	15.70	13.31	12.72	13.64
Ранг	**17.00**	**21.00**	**32.00**	**14.00**	**21.00**
Цвет	Место 1	Место 2	Место 3	Место 4	Место 5

Таблица 1. Ошибки классификаторов (в %).

в матрице D^{-1}. Соответственно, в ходе обучения решается задача одномерной оптимизации $\mathbf{w}_{MP} = \theta\mathbf{u} = \arg\max_{\theta} \left(L(T|X, \theta\mathbf{u} + d^{-1}\theta^2)\right)$.

Заметим, что в отличие от метода релевантных векторов, в предложенном подходе не требуется итеративного подбора коэффициентов регуляризации, а формула для оптимального регуляризатора сразу выписывается в явном виде.

Эксперименты

В таблице 1 представлены результаты экспериментов, проведенные на реальных задачах из репозитория UCI [4]. Сравнение проведено с классическим (RVM) и вариационным (VRVM) [2] методом релевантных векторов, методом опорных векторов (SVM) и логистической регрессии (LogReg). Отдельно можно отметить, что скорость обучения практически не отличается от скорости обучения логистической регрессии и значительно быстрее обучения метода релевантных векторов.

Работа выполнена при поддержке РФФИ, проекты № 06-01-08045, № 05-07-90333, № 06-01-00492 и № 07-01-00211.

Литература

[1] *Kropotov D. A., Vetrov D. P.* Optimal Bayesian Linear Classifier with Arbitrary Gaussian Regularizer // 7th Open German-Russian Workshop on Pattern Recognition and Image Understanding (OGRW2007), Ettlingen, 2007.

[2] *Bishop C. M., Tipping M. E.* Variational Relevance Vector Machines // Uncertainty in artificial intelligence (UAI-2000), 2000— P. 46–53.

[3] *Tipping M. E.* Sparse Bayesian Learning and the Relevance Vector Machine // Journal of Machine Learning Research. — 2001. — Vol. 1, № 5. — P. 211–244.

[4] *Asuncion A., Newman D. J.* UCI (Machine Learning Repository) — 2007. — `www.ics.uci.edu/~mlearn/MLRepository.html`.