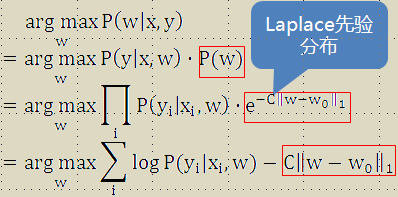
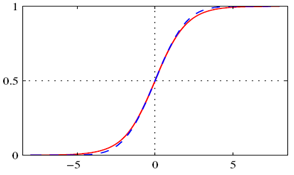
Logistic Regression

Logistic regression（ [LR](http://docs.babel.baidu.com/doc/bbcbe063-4b85-4277-9706-3f1a65fb8e04)）可以看作是基于MAP（maximum a posteriori）估计的优化问题。利用Bayes公式，假设特征weight（w）服从Laplace分布（式中x表示样本特征，y表示label）：



将likelihood表示为sigmoid函数（logistic函数，如下面的图）形式，令w0=0（假设特征weight分布以0为中心，提高泛化性），就得到了logistic regression的优化公式。Sigmoid函数的优点是，自然归一化到（0，1），且连续可导，导数容易计算。Likelihood表示为：

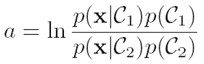
http://wiki.baidu.com/download/attachments/86030881/image2012-6-1414%3A1%3A40.png?version=1&modificationDate=1429931254560&api=v2   http://wiki.baidu.com/download/attachments/86030881/image2012-6-1414%3A1%3A48.png?version=1&modificationDate=1429931254560&api=v2



关于一个条件概率为何可以表示为sigmoid函数形式，可以通过Bayes公式理解：

http://wiki.baidu.com/download/attachments/86030881/image2012-6-1414%3A19%3A45.png?version=1&modificationDate=1429931254561&api=v2

式中



对于logistic regression，另一种观点是，可以看作最大似然估计（maximum likelihood estimation, MLE）。后面的先验项是正则项，以提升模型泛化性。

理论上，LR能够求得全局最优解（两个凸函数的和仍为凸函数）。可以通过OWLQN（Orthant-Wise Limited-memory Quasi-Newton）算法求解（L1范数0点不可导）。选择L1范数的好处是：可以得到更稀疏的weight向量，从而具有更强的特征选择能力，以及便于计算和存储。这是由于L1范数象限内导数（将weight推向0的作用力）恒定。对比起来，L2范数越趋近0，导数绝对值越小（作用力越弱），只能得到很多接近0（但不是0）的weight。

Prior Model

在前面logistic regression的推导公式中，prior信息通过w0引入到模型中，表示特征weight服从以w0为中心的Laplace分布。此时w0即为prior weight，通过其它数据，或是已有规则、知识得到。

使用prior model时，需要考虑引入的prior weight与当前特征、数据是否match。如果不match，效果可能不理想，甚至变差。需要根据具体情况，设计引入哪些特征的prior，哪些特征prior weight不可改变，哪些可以改变……

这里仅对相关知识做简单介绍，更详细内容请参考网上信息或model组资料。