jasonfreak
一个懒惰的人，总是想设计更智能的程序来避免做重复性工作

关于线性模型你可能还不知道的二三事（二、也谈民主）

系列博文

- 关于线性模型你可能还不知道的二三事（一、样本）
- 关于线性模型你可能还不知道的二三事（二、也谈民主）
- 关于线性模型你可能还不知道的二三事（三、特征值与奇异值的魔力）

目录

---

1 如何更新权值向量？

在关于线性模型你可能还不知道的二三事（一、样本）中我已提到如何由线性模型产生样本，在此前提下，使用不同机器学习算法来解决回归问题的本质都是求解该线性模型的权值向量W。同时，我们常使用线性的方式来解决分类问题：求解分隔不同类别个体的超平面的法向量W。不论回归还是分类，都是求解向量W，而求解的核心思想也英雄所见略同：向量W倾向于指向某些“重要”的个体。然而哪些个体是重要的呢？不同的机器学习算法有不同的定义。

---

2 最小均方法（LMS）与感知机：低效的民主

最小均方法（LMS）使用的随机梯度下降法与感知机的训练法则类似，两者都是迭代更新的方式。假设本次迭代中的权值为W，那么更新后的权值W'为（eta为更新率）：

机器学习(2)
数据分析(2)
数据挖掘(8)
特征工程(2)

随笔档案

2016年11月 (1)
2016年7月 (3)
2016年6月 (4)
2016年5月 (2)
2016年4月 (2)

最新评论

1. Re:使用sklearn做单机特征工程
标准化是依照特征矩阵的列处理数据，归一化是依照特征矩阵的行处理数据，这个不太理解，博主可以解释下吗？
--Stone1111
2. Re:使用sklearn进行集成学习——理论
你好博主！我想问一下“在bagging和boosting框架中，通过计算基模型的期望和方差，我们可以得到模型整体的期望和方差。为了简化模型，我们假设基模型的权重、方差及两两间的相关系数相等。”这里接.....
--implus
3. Re:使用Python进行描述性统计
很好，很清晰，赞！
--iuwai
4. Re:使用sklearn优雅地进行数据挖掘
@会飞的蜗牛引用@魔灵幽亭在你的数据集DataFrame上加一句df = df.fillna(0)...
--liuer2009
5. Re:谁动了我的特征？——sklearn特征转换行为全记录
@hnxsm这个调的哪里的...
--hustenn

阅读排行榜

1. 使用sklearn做单机特征工程(47866)
2. 使用sklearn优雅地进行数据挖掘

随机梯度下降法：

$$W' = W + \eta * (y_j - W * X_j) * X_j$$

感知机：

$$W' = W + \eta * \left(y_j - sgn(W * X_j)\right) * X_j$$

通过观察可知，权值更新是一个迭代的过程，不论是回归（最小均方法）还是分类（感知机），权值更新时视当前轮次中误差大的个体为“重要”的个体。这种权值更新办法比较直观，但是同时也比较低效：人人都有发言的权利，每次只考虑部分人，容易顾此失彼。

---

3 最小二乘法：完美的民主

二乘即是平方，最小二乘法旨在于求解权值向量W使得误差平方和最小：

$$\min_W \frac{1}{2} * \sum_j^m (y_j - W * X_j)$$

通过对权值向量的每个分量进行求导可得：

$$W = Y * X^T * (X * X^T)^{-1}$$

至此，我们可以发现最小二乘法可解的条件为特征矩阵X是可逆的。假设特征矩阵X的样本容量n=m，那么上式进一步化简得：

$$W = Y * X^{-1}$$

使用求解出来的权值向量W'对未知个体x'进行预测，本质就是计算：

$$y' = W * x' = Y * X^{-1} * x'$$

在《关于线性模型你可能还不知道的二三事（一、样本）》中我们已经揭开了特征矩阵X的逆矩阵的意义，因此以上的计算过程可以概括为：首先使用X的逆矩阵乘以未知个体x'，得到可以准确描述未知个体x'与特征矩阵X中已知个体相似度的列向量，然后以此为基础，使用加权求和的方法来计算未知个体x'的目标值。

到此，最小二乘法所诠释的完美民主已显见：在每个人都不能由其他人代表的前提下，看未知的个体与谁更相似，那么目标值也与之更相似。

没错，之前我们假设了特征矩阵X的样本容量n=m，但是大多数情况下n是大于m的。这种情况下权值向量计算公式无法进一步化简。同样在《关于线性模型你可能还不知道的二三事（一、样本）》中我们提到，可以转化原问题为：

(42921)
3. 使用Python进行描述性统计(34743)
4. 使用sklearn进行集成学习——理论(22104)
5. 使用sklearn进行集成学习——实践(22084)

评论排行榜

1. 使用sklearn优雅地进行数据挖掘(21)
2. 使用sklearn做单机特征工程(18)
3. 使用sklearn进行集成学习——理论(9)
4. 虎扑论坛装备区到底有没有李宁水军？——论坛水军发现实践(6)
5. 使用sklearn进行集成学习——实践(3)

推荐排行榜

1. 使用sklearn做单机特征工程(20)
2. 使用sklearn优雅地进行数据挖掘(17)
3. 使用sklearn进行集成学习——理论(8)
4. 使用Python进行描述性统计(6)
5. easyconf——基于AugularJS的配置管理系统开发框架(5)

$$Y * X^T = W * X * X^T$$

这时，我们可以设新的特征矩阵X'和新的目标值向量Y'为：

$$Y' = Y * X^T$$
$$X' = X * X^T$$

到此，新的特征矩阵X'是m×m的方阵，可以求其逆矩阵了（当然，这还是在原特征矩阵的秩等于m的前提下）。因此有：

$$W = Y * X^T * (X * X^T)^{-1}$$
$$= Y' * X'^{-1}$$

不难看到，上式同样也是诠释了完美的民主，只是特征矩阵X变成了X'，目标值向量Y变成了Y'而已。

---

4 支持向量机：现实的民主

完美的民主可遇而不可求，如果特征矩阵X的秩小于m呢？此时最小二乘法便不奏效了。我们期望无论特征矩阵X的秩是否小于m，仍然可以高效地求解权值向量W。

我们可以利用支持向量机解决该问题。不妨直接看到权值向量的最终结果（具体推导可参考《支持向量机通俗导论（理解SVM的三层境界）》）：

$$W = \sum_{j}^{n} \alpha_j * y_j * X_j^T$$

使用上式计算出来的权值向量W对未知个体x'进行预测的原理是显见的：首先将未知个体与特征矩阵X中的个体相乘得到对应的相似度，然后以此相似度乘以alpha的分量，最后在此基础上以加权求和的方法来计算未知个体x'的目标值。然而，alpha到底是什么呢？

对支持向量机有一定了解的同学肯定会有一个基本的认识：支持向量为alpha分量不为0的点，该点位于间隔边界上。也就是说，最终的权值只会考虑作为支持向量的样本！然而，进一步，很少有人会去思考：间隔边界上的点都是支持向量吗？支持向量所对应的alpha的分量值大小服从什么规律吗？支持向量为什么叫支持向量呢？这些问题暂且不表，在支持讲支持向量机时进行进一步分析。

此时，我们可以引出结论：支持向量机代表的是一种现实的民主，我国的人民代表大会制也是如此。

---

5 总结

这次，我们探讨了3种常见的线性模型权值向量求解思路。从LMS和随机梯度下降到最小二乘，再到支持向量机，人们求解自然科学问题的思路与求解社会科学问题的思路走到了一起。最近的一件小事带给我启发：居住的小区需要对某一些问题进行决策，一开始由热心居民每家每户听取意见，结果迟迟拿不定主意，越听越糊涂。到最后，只好选出业主委员会，由业主委员会代表各个特色群体，问题才得以解决。

之前对线性模型的权值求解过程和结果都“记得”非常熟悉，但是其真正意义（特别是最小二乘）没有去深究。而这次能够受到启发，并且联系到现实生活中，也算是对线性模型有了更进一步的认识吧。

---

6 参考资料

1. 支持向量机通俗导论（理解SVM的三层境界）
2. sklearn svc model

分类: 数据挖掘

标签: 数据挖掘, 线性模型

« 上一篇：关于线性模型你可能还不知道的二三事（一、样本）
» 下一篇：关于线性模型你可能还不知道的二三事（三、特征值与奇异值的魔力）

posted on 2016-06-16 17:27 jasonfreak 阅读(1711) 评论(0) 编辑 收藏

最新IT新闻:

· Rocket Lab成功发射第一颗卫星
· 淘宝卧榻之侧，岂容拼多多安睡？
· Docker日志的10大陷阱
· OpenSSL改变开发策略：转用GitHub issue讨论补丁
· 深圳回应制造业外迁 市长称不会出现产业空心化
» 更多新闻...

最新知识库文章:

· 领域驱动设计在互联网业务开发中的实践
· 步入云计算
· 以操作系统的角度述说线程与进程
· 软件测试转型之路
· 门内门外看招聘
» 更多知识库文章...