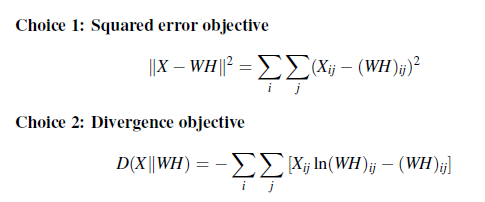
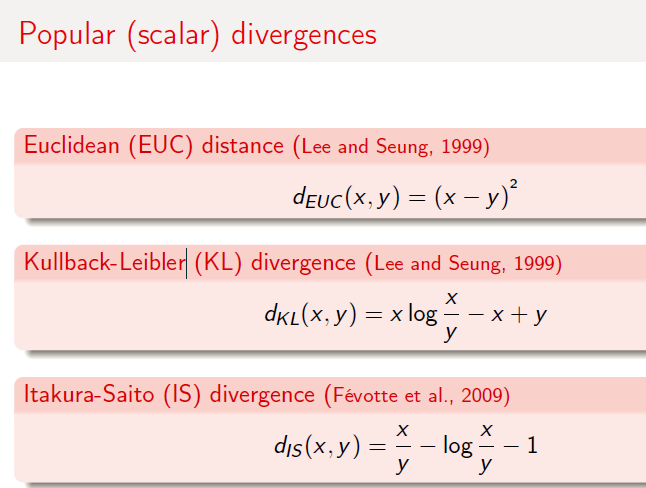
Some thoughts

1, NMF model 主要包括2类model

* Squared error objective
* Divergence objective, 统计解释 Maximum Likehood, xij ~ Pois((WH)ij)



其他的我看到有



请问我们是准备用哪些model呢？

2.LDA 就是双重Direchlet函数加贝叶斯框架，FICO文章把问题转化为了LDA模型，还是很有创新性的，一个交易主体就是一篇文章，一个主题所产生的每一笔交易就是文章中的一个字，这个字由5个feature形成。因为一笔交易就是一个字，不需要考虑2阶，3阶，这个字里面的字母就跟普通单词里面的字母一样是没有什么含义的，不同的组合形成不同的字，这里面并不存在regression里面交互项的概念。因为他把整个问题都转化成了LDA 模型，设计也是非常合理的。

3.如果我们要用NMF,从以下几个方面考虑：

* LDA属于生成模型，对P(x,y)先做了模型假设，假设Direchlet分布，这种假设太强，比较具有局限性，完全可以用NMF这种判别式方法替代。由于交易的产生用poison model特别合适，用NMF的Divergence model也很好，背后能有直观的统计解释
* FICO把5个feature连在一起成一个字，没有考虑到2个feature或者3个feature组合起来形成的关系，如果我们用NMF的话, 就也许能考虑交互项关系了，把这些组合全部列出来放到NMF里面run? 我们有结果出来了吗？