

データマイニング

第1回

ホームページ

- ◆ 今年の東日本大震災で、サーバのハードディスクがクラッシュしまして
- ◆ せっかくなので引っ越ししました

<http://stat.inf.uec.ac.jp/dokuwiki/>

<http://appl.stat.inf.uec.ac.jp/>

データマイニングとは

- ◆ データを分析すること（ビジネス）
- ◆ 機械学習論をデータ分析に応用すること、統計的データ解析には含まれない手法も用いること（研究）

データマイニング

- ◆ 統計的データマイニングと非統計的データマイニングがある
- ◆ 多変量解析と学習理論のはざま
- ◆ 発見の再現性より発見そのものを重視

類似の言葉

- ◆ データ分析、データ解析
- ◆ 統計解析、多変量解析
- ◆ パターン認識
- ◆ 機械学習

いずれにせよ

- ◆ データから何かを学ぶ手法
(こちらは学術的)
- ◆ またその適用プロセス
(こちらの方がお金になる)

この講義について

- ◆ 統計学の復習(さらっと最尤法など)
- ◆ 統計的学習理論(リスク最小化問題)

教師あり学習、教師なし学習

この講義について

- ◆ 線形学習機械(判別分析、ロジスティック回帰)
- ◆ 非線形学習機械(決定木、サポートベクトルマシン、ブースティング)

統計的データマイニング？

- ◆ 再現性を重視するアプローチ
- ◆ ちょっとうるさい

非統計的データマイニング？

- ◆ 発見に特化する流儀
- ◆ 結果重視
- ◆ かなり自由
- ◆ 統計家の前では微妙

この授業の寿命

- ◆ 今年限りです。
- ◆ 来年度は「離散数学」を受けて、単位をとってもらいます。

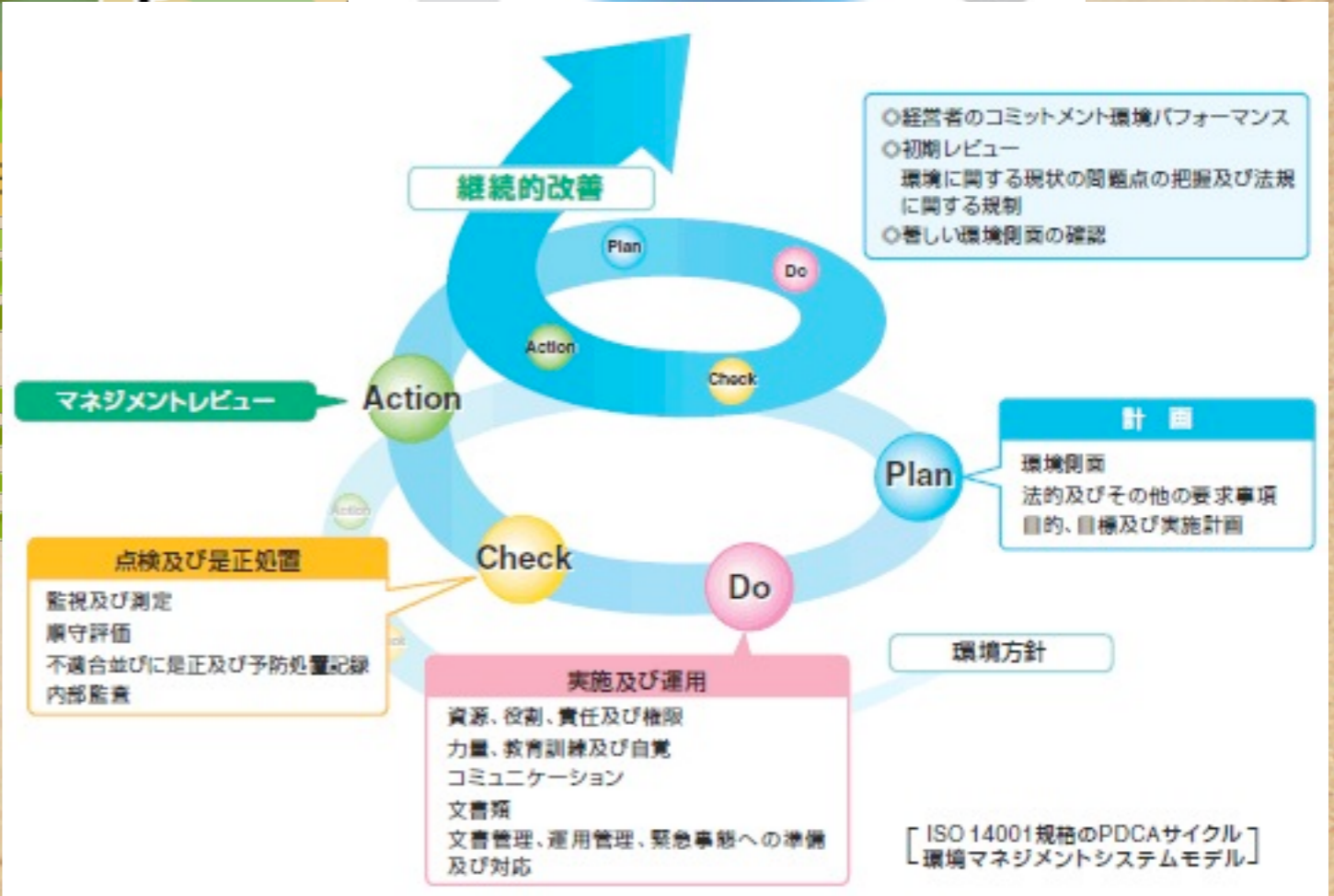
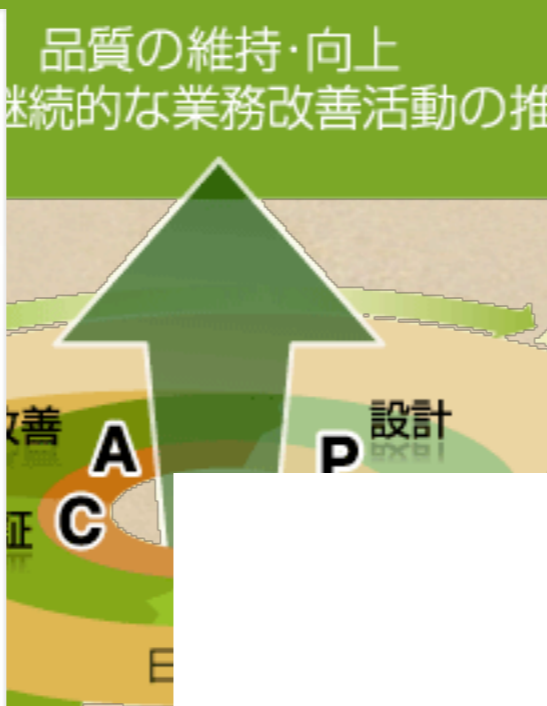
離散数学って？

- ◆ 離散の数学、です

なんで離散数学なの？

- ◆ そこに離散数学があったから？

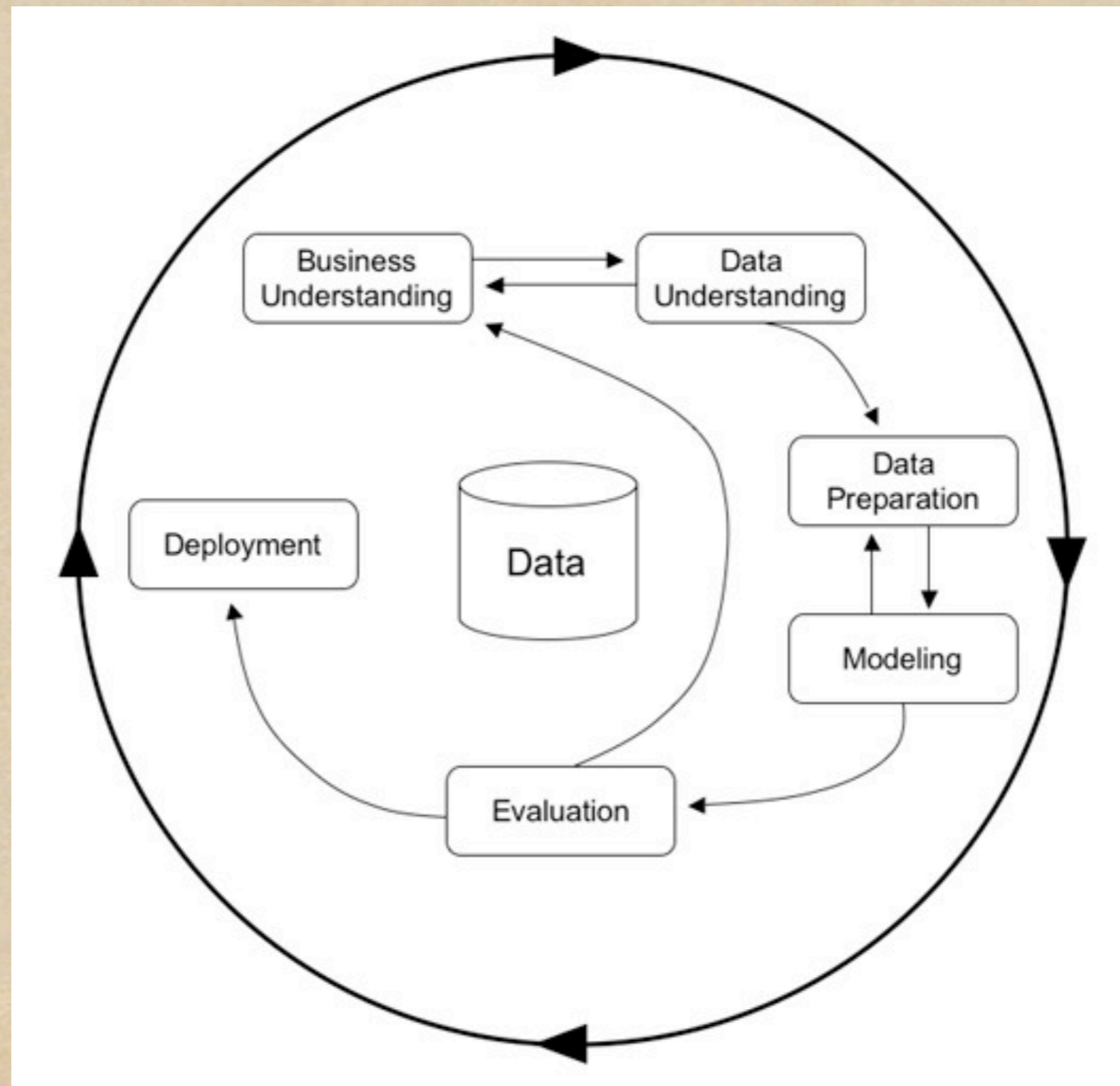
PDCA



一回で終わるな、と

- ◆ 物事、試みてみただけ(やりっぱなし)ではなく、その結果を評価し、工夫して改善していくことが重要
- ◆ データの分析(データマイニング)もそう

プロセスとしてのDM



R言語

- ◆ 対話的なデータマイニング環境
- ◆ 分析プロセスをコードにすべて残せる

履修に際してのお願い

- ◆ R言語が使える環境を用意してください
- ◆ レポートはWordファイル形式

成績評価

- ◆ 3～4通のレポート課題（最近は、返却してません）
- ◆ ひとつのデータセットについて
 - ◆ データの要約
 - ◆ 線形学習機械による分析
 - ◆ 非線形学習機械の適用

忘れないうちに

- ◆ T科の人で、来年度に卒業研究をしたい人、12月8日は午後から夜まで、開けておいてください
- ◆ H科もそのあたりか調布祭のはず