主成分分析についてまとめてください。（自分が理解できていることを採点者に伝えてください。）

[求め方]
１、(共分散=Multiがなるべく大きくなるような)共分散行列を求める。
２、固有値・固有ベクトルを求める。
→やってることは写像を求めているだけ

[説明]
機械学習において主成分分析は異常検知としても役立っています。そちらに寄せて説明します。
データ分析の種類の一つとして、教師なし学習による異常検知というものがあります。
ほとんどが正常なデータでまれに異常なデータが混じっている、
その異常発生のパターンや異常と他の要因との紐付きがいまいちつかみきれていないというような場合、
教師あり学習による2値分類がうまくワークしない、といった状況がありえます。
そういった場合には、正常パターンを教師なし学習で学び、その正常パターンから外れているものを異常とする、という方法が有効です。
この異常検知の方法の一つとして、主成分分析（PCA）をベースにした異常検知の方法があるということです。
そして主成分分析とは、"与えられたデータ情報の次元を圧縮する手法"の一つである。
またPCAは分散の値が大きいほどよい。なぜなら、バラついているということは、
各データ点１つ１つの違いをより多く情報として保っていることになるからである。
逆に言えば、バラツキの少ない方向というのは、
各データが共通して持っている自明な情報なので削除してしまっても問題ないのである。
まとめると、主成分分析では、データの次元圧縮を行う。
その際、圧縮後のデータの分散が大きくなるような射影をすることで、特徴量を自動的に抽出するのである。

主成分分析について素人にも分かるように簡潔に説明してください。

主成分分析とは、"与えられたデータ情報の次元を圧縮する手法"の一つである。
例えば、肥満度を測るためのBMIという指標が存在する。BMIは体重÷(身長)^2という計算式で導出されるものであるが、
よく考えてみると元のデータは(身長,体重)という2成分を持つデータであったのに対し、BMIはただ1成分の数値となっている。
これが次元の圧縮であり、情報を削ぎ落としたにも関わらず、肥満度という特徴を表すのに十分な情報を持っている。
このように、データに適切な処理を行えば、情報量の削減と特徴の抽出を同時に行う事ができる。

主成分分析について数式を用いて説明してください。

また、以下のキーワードを用いて説明すること

- 共分散行列
- 固有値問題

「分散」とは、ある一次元のデータ
$((x_1, y_1), (x_2, y_2), (x_3, y_3))$ に対して、平均 $\overline{x}$ または $\overline{y}$ からの差、の積平均にあたる。式で書けば次のようになる。

$$\sigma_{xx} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2}{3}$$

「共分散」とは、ある多次元のデータ $((x_1, y_1), (x_2, y_2), (x_3, y_3))$ に対して、平均 $\overline{x}$ または $\overline{y}$ からの差、の積平均にあたる。式で書けば次のようになる。

$$\sigma_{xy} = \frac{(x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y}) + (x_3 - \overline{x})(y_3 - \overline{y})}{3}$$

分散は式や定義から、データのバラつきを表していることが直感的にわかる。また、共分散はすぐにはわかりづらいが、xデータとyデータの相関度合いを示すものである。続いて、これらを要素として並べた「共分散行列」なるものが存在する。

$$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} (x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 & (x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y}) + (x_3 - \overline{x})(y_3 - \overline{y}) \\ (x_1 - \overline{x})(y_1 - \overline{y}) + (x_2 - \overline{x})(y_2 - \overline{y}) + (x_3 - \overline{x})(y_3 - \overline{y}) & (y_1 - \overline{y})^2 + (y_2 - \overline{y})^2 + (y_3 - \overline{y})^2 \end{pmatrix}$$

これと、何らかの行列を与えられた時に固有値と固有ベクトルを求めるという「固有値問題」があるが、 共分散行列から固有値・固有ベクトルを求める。
その固有値・固有ベクトルを用いて、データと主成分の内積をとると
主成分分析の求める主成分が求まる。

主成分分析をPythonで実装してください。

別ファイルを参照ください。

その他、今回の授業で学んだことを記述してください。

理論を学んだだけで、実際に分析を行った訳ではない為、
今後グラフ化などされた物を見たり、実際に分析を行った時に役立つことを
とても期待しています。