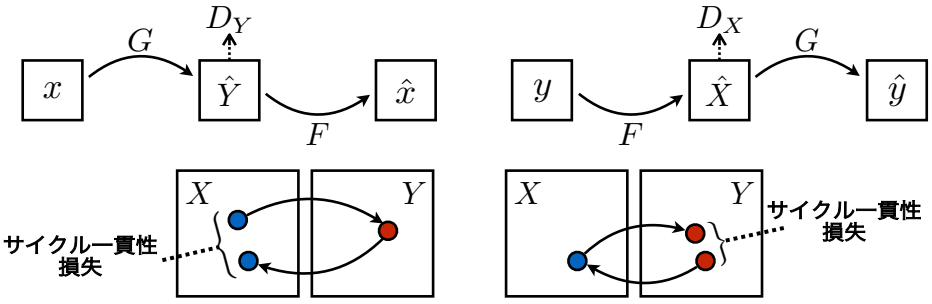


図18: サイクル GAN の概念図 [80] Fig.3 を改変

2.2 VAE

変分自己符号化器 **VAE**[*26] はキングマとウェリング[33] とレゼンデら[58] によって 2014 年に提案された変分ベイズに基づく 半教師ありネットワーク です。図19 に通常の自己符号化器と変分自己符号化器 (VAE) の相違を示しました。自己符号化器は，中間層での結合が潜在変数のサンプリングに置き換えられています。潜在変数である未知の変数の確率分布を類推する手法にはいくつか存在します。このうちベイズ推論を援用して推定を行うため変分ベイズとも呼ばれる手法を変分自己符号化器 VAE と呼びます。 VAE はデータ x が与えられたときモデルを記述する潜在変数 z を類推するための符号化器，復号化器に DNN を用いるモデルです。

VAE の特徴を列挙すれば，以下のようになります:

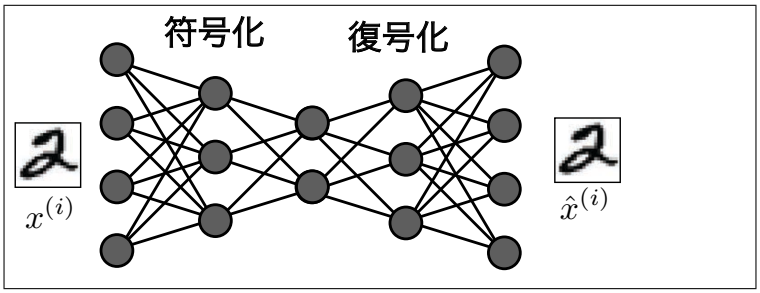
1. 符号化器と復号化器とを持つエンコーダ-デコーダモデルである
2. 符号化器と復号化器とは共に DNN が用いられる
3. 符号化器と復号化器との間に潜在変数 z を仮定し，z の確率密度を推定する
4. 符号化器はデータ x を潜在変数 z の空間へ写像する
5. 復号化器は潜在変数 z の確率密度をベイズ推論によってデータ x を再構築する生成モデル である
6. VAE の学習時に必要となる目的関数では 再パラメータ化トリック[*27] を用いることで勾配降下法が利用できる

変分法とは歴史的には物理学でニュートンの運動法的式を一般化する過程で発展した一般的手法です。ディープラーニングとの関連では，最適化手法や確率アルゴリズムの基盤となる考え方です。VAE ではベイズ推論を用いて変分法による解を求めます。このために 変分ベイズ[*28] または 変分推論[*29] と呼ばれます。変分ベイズ法の解法には，平均場近似が主として用いられてきましたが，キングマらの提案手法により応用が広がりました。

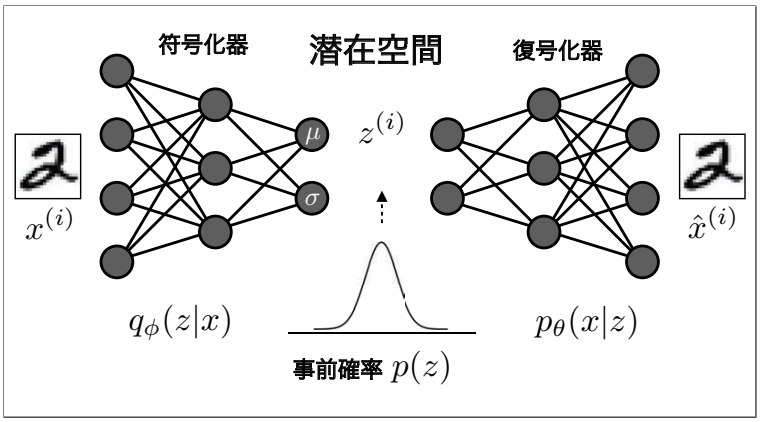
図19bに即して，やや詳しく書くと次のようになります。 VAE は観測可能 (で複雑) なデータ x を経験分布 $q_\phi(x)$ と，比較的単純な分布である潜在変数 z の空間との間の確率的にマッピングすることを学習します。 生成モデル p_θ は x と z との同時分布 $p_\theta(x,z)$ を学習します。この同時分布を因子分解すると $p_\theta(x,z) = p_\theta(z)\,p_\theta(x\,|z)$ となります。この式は，観察可能な変数 x と潜在変数 z の同時分布が，潜在変数 z の事前分布と，確率的に振る舞う復号化器の出力 $p_\theta(x\,|z)$ の積であることを示しています。 推論モデル[*30] q_ϕ は，確率的な符号化器 $q_\phi(z\,|x)$ であり，生成モデルの出力である事後分布 $p_\theta(z\,|x)$ を近似します。それぞれの記号の対応関係を図19b で確認してください。ϕ, θ

[*26] variational auto-encoders
[*27] reprametrization trick
[*28] variational Bayes
[*29] variational inference
[*30] inference model

は原著論文[33] の記号に対応していますが，慣れないうちは無視して考えてください。



(a) 自己符号化器の模式図



(b) 変分自己符号化器の模式図

図19: 自己符号化器(a)と変分自己符号化器(b)の違い [65] を改変

潜在変数 z の真の分布 $q(z)$ と入力 x が与えられたときの z の推定値を $p(z|x)$ とします。そうすると VAE の目標関数は両分布の **KL ダイバージェンス**を小さくすることが求められます。 KL ダイバージェンスは $\mathrm{KL}(X||Y)$ あるいは $D_{\mathrm{KL}}(X||Y)$ などと表記されます。 KL ダイバージェンスは X と Y との分布間の乖離の度合い，ある種の距離と考えることも可能です。しかし，KL ダイバージェンスは，どちらか一方から他方との乖離 (距離)を考えるかによってその値が異なります。 $\mathrm{KL}(X||Y) \neq \mathrm{KL}(Y||X)$ 等号が成立するのは，両分布が等しい場合に限られます。

VAE では，この KL ダイバージェンスを直接計算することが難しいため， KL ダイバージェンスを計算することと等価な ELBO[*31] を最小化することが行われます。 ELBO は 変分下限とも呼びます。

入力データ x を用いたモデルの対数尤度を $\log(x)$ とすれば，この対数尤度を最小化するために潜在変数 z を導入し，導入した z を含めた関数の最適化とみなすことができます。このため単純な最適化ではなくなり，汎関数の最適化となります。これを 変分法 (variational method) と呼びます。変分法は，物理の分野で開発された手法ですが，広く条件付き最適化に応用可能で，経済学，社会学，統計学，でも用いられています。

VAE の模式図を図20に示しました。図20は左から入力が与えられます。入力データは推論モデルを通して潜在空間上の z を予測します。実際にはサンプリングがおこなわれ，さらに z と入力データに基づいて生成モデルから目的関数が計算されます。

[*31] Evidence Lower BOund

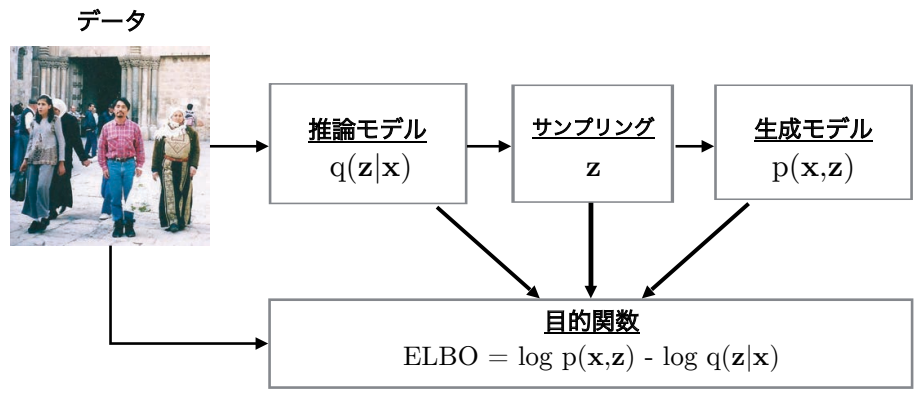


図20: 変分自己符号化器の目標関数 ELBO の概念図 [34] Fig. 2.2 を改変

このことは，GAN と良い対比になります。

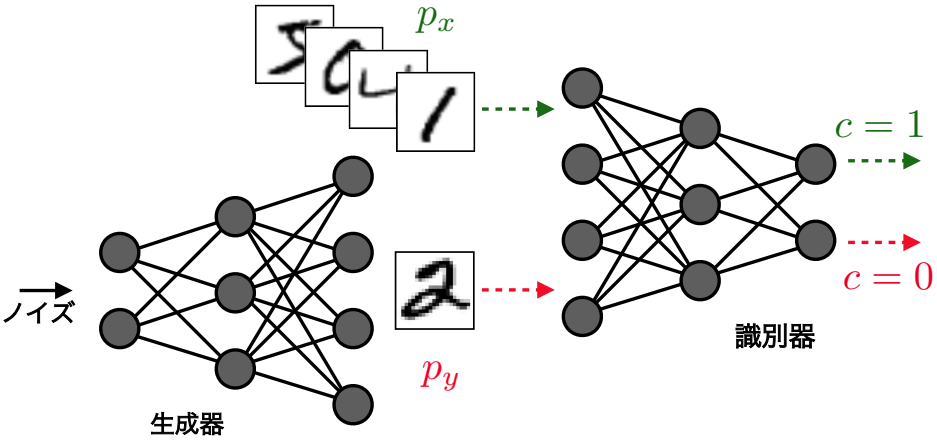
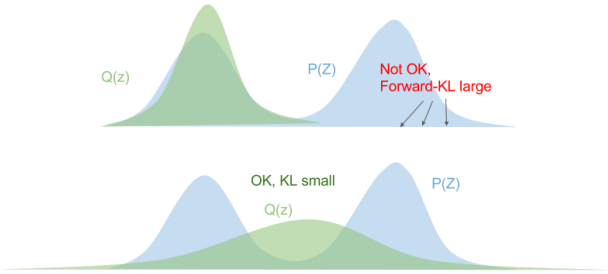


図21: GAN では生成モデルの訓練に敵対的密度比推定を使用する。これは 2人ゲームと見ることができる。識別器はサンプルが実在か，または生成器が生成したデータかを予測しようとする。生成器は実在サンプルの分布を模倣することで識別器を欺こうとします。

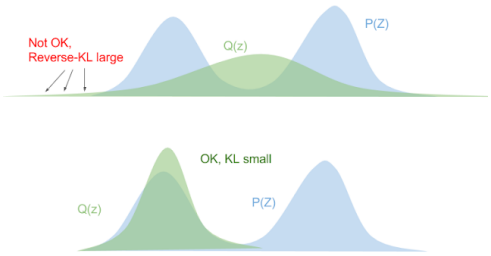
■ELBO　VAE の損失関数としては ELBO[*32] を用います。 ELBO 変分下限と呼ばれていたものです。

KL ダイバージェンスは，2 つの分布間の距離に相当する量を与えます。ですが KL ダイバージェンスは通常の距離と異なり非対称で，どちらの分布を基準に考えるかによって値が異なります。すなわち $\mathrm{KL}(P\|Q) \neq \mathrm{KL}(Q\|P)$ です。下図22 にその関係を示しました。青い曲線は真の事後分布とします，例えば双峰性の分布であるとします。緑の分布は最適化を介して青い密度に適合させる変分近似による分布を表すものとします。これを フォワードKL と呼びます。図22 右のように，双峰性の真の分布を単峰性の分布で近似することを考えます。このとき，一方の峰に当てはまるように調整すると，もう一方の峰の値についての当てはまりが悪くなり結果として右下図のような裾野の広い分布を得ることになります。反対に，緑の単峰性の分布を青の双峰性の分布で近似しようとする リバースKL を考えます。このとき基準となる真の分布である単峰性の分布の確率密度がほとんど 0 の領域では，推定する分布がどのような値を取ろうとも KLダイバージェンスの値に影響を与えないため，いずれか一方の峰が真の分布と重なるような値を得ることになります。

[*32] Evidence Lower BOund



(a) フォワードKL



(b) リバースKL

図22: KLダイバージェンスの非対称性 A Beginner's Guide to Variational Methods: Mean-Field Approximation より

このようなフォワード，リバース KL ダイバージェンスの値から VAE の表現性能などを考えることが可能です。

■再パラメータ化トリック　確率変数 z はデータ x と変分パラメータとから計算されます。ところが変分パラメータを直接微分することが難しいため，誤差逆伝播法を用いた学習を行う場合に工夫が必要になります。簡単に述べると z を平均 μ 分散 σ^2 に従う正規分布だとみなしてしまうことで $q_\phi(z|x) = \mathcal{N}\left(z; \mu, \sigma^2\right)$ と考えることにします。データ x が与えられたとき符号化器の出力を μ と σ と誤差 ϵ とみなして潜在変数 z を近似することを 再パラメータ化トリック[*33] と呼びます。潜在変数 z は多次元で複雑なはずですが，各次元が独立であると仮定すれば計算が簡単になります。さらに z の各次元の分散 σ^2 が等しいという仮定をおけば，求めるパラメータ数が減り計算が簡単になります。

本来は複雑であるはずの潜在変数を，最パラメータ化トリックによってどこまで簡略化できるかについては慎重になるべきです。ですが，仮にこのような簡略化が可能であれば，潜在変数の解釈可能性，表現可能性を議論できます。これにより DNN の解釈可能性が広がると考えられます。このような動向に基づきベータVAE[6, 24]，因子化VAE[32]，トータル相関 VAE[8] などが提案されています。

2.2.1　解きほぐし表現

解きほぐし表現 とは disentaglement あるいは disentangled represenation の訳です。日本語としての定訳がないためここでは仮に解きほぐしとしました。解きほぐし表現とは，畳み込み多層ニューラルネットワークによって得られた表現が，解釈不能であった場合，より簡易で単純な表現を与えることを指します。大量かつ多次元データを取り扱い，かつ人間に比肩しうる精度を達成してきたディープラーニングモデルへの要求として次に求められる課題として，如何に解釈可能な表現を提示できるかという点が挙げられます。

変分自己符号器 VAE は，入力情報を忠実に再現する 雑音除去自己符号器[*34] と異なります。符号器と復号器との間に潜在変数を仮定し，その潜在変数の分布をベイズ推論によって変分推定することに特徴があります。再パラメータ化トリックの項でも説明したように，潜在変数間に直交性を仮定すれば，解釈可能な表現を得ることが期待できます。このことに端を発して多くの研究がなされています。

3　自然言語処理

単語の意味を表現できるベクトル空間モデル（word2vec）とニューラル画像脚注付け、ニューラルチューリングマシンを取り上げます。画像処理と自然言語処理の融合による知的情報処理を概観します。

[*33] reparametrization trick
[*34] denoising autoencoders[71]