β VAE

Weronika Michoń

24 marca 2020

W pracy postaram się przybliżyć czym są i do czego służą β wariacyjne autoenkodery (w skrócie β VAE).

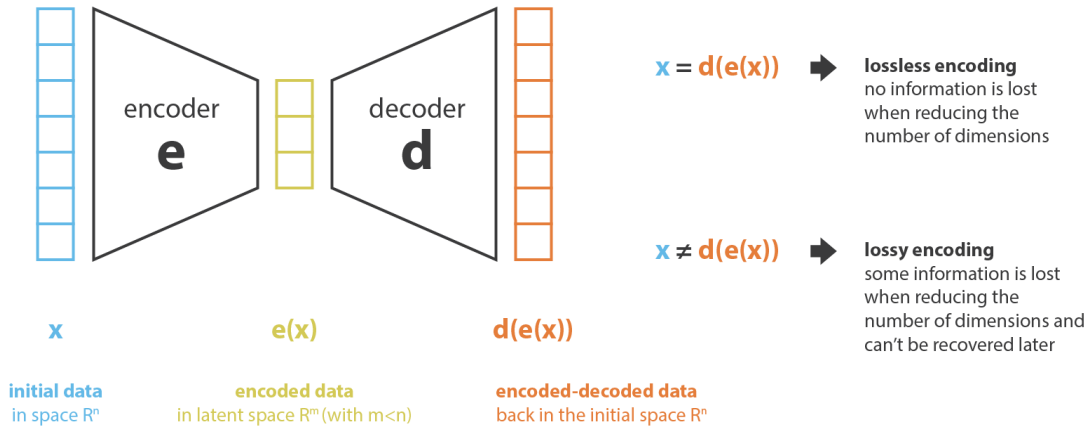
Zanim przejdziemy do ogólnego przedstawienia β wariacyjnych autoenkoderów, zacznijmy od tego co to jest autokoder. Autokodery to rodzina modeli sieci neuronowych, których celem jest poznanie skompresowanych ukrytych zmiennych danych wielowymiarowych.

Łatwiej będzie zrozumieć β wariacyjne autoenkodery, gdy przejdziemy przez autoenkodery od początku. Plan pracy przedstawia się następująco:

- Co to jest autoenkoder?
- Co to jest przestrzeń ukryta (latent space) i dlaczego ją regulujemy?
- Co to jest VAE?
- Jak generować nowe dane wykorzystując VAE?
- Co to jest β VAE?

1 Redukcja wymiarów

W uczeniu maszynowym redukcja wymiarów jest procesem zmniejszania liczby cech opisujących dane.



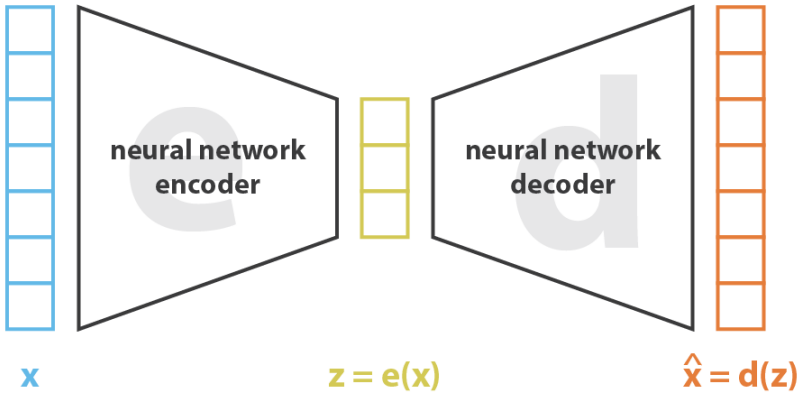
Rysunek 1: Ilustracja zasady redukcji wymiarów za pomocą enkodera i dekodera.

Najpierw zdefiniujmy encoder czyli proces, który tworzy reprezentację „nowych cech” z reprezentacji „starych cech” (przez wybór lub ekstrakcję), natomiast decoder to proces odwrotny. Redukcję wymiarów można zatem interpretować jako kompresję danych, w której encoder kompresuje dane (z przestrzeni początkowej do przestrzeni kodowanej, zwanej również przestrzenią ukrytą), podczas gdy dekoder dekompresuje je. Oczywiście, w zależności od początkowej dystrybucji danych, wymiaru ukrytej przestrzeni i definicji enkodera, kompresja ta może być stratna, co oznacza, że część informacji jest tracona podczas procesu kodowania i nie można jej odzyskać podczas dekodowania.

Głównym celem metody redukcji wymiarów jest znalezienie najlepszej pary enkoderów / dekoderów w danej rodzinie. Innymi słowy, dla danego zestawu możliwych koderów i dekoderów szukamy pary, która zachowuje maksimum informacji podczas kodowania, a zatem ma minimalny błąd rekonstrukcji podczas dekodowania.

2 Autokoder

Omówmy teraz autokodery i zobaczmy, jak możemy wykorzystać sieci neuronowe do redukcji wymiaru. Ogólna koncepcja autokoderów jest dość prosta i polega na ustawieniu enkodera i dekodera jako sieci neuronowych oraz nauce najlepszego schematu kodowania-dekodowania przy użyciu iteracyjnego procesu optymalizacji. Tak więc przy każdej iteracji zasilamy architekturę autoencodera pewnymi danymi, porównujemy dane wyjściowe kodowane i dekodowane z danymi początkowymi i propagujemy błąd w architekturze z powrotem, aby zaktualizować wagi sieci.



Rysunek 2: Ilustracja autoencodera z funkcją straty (loss function).

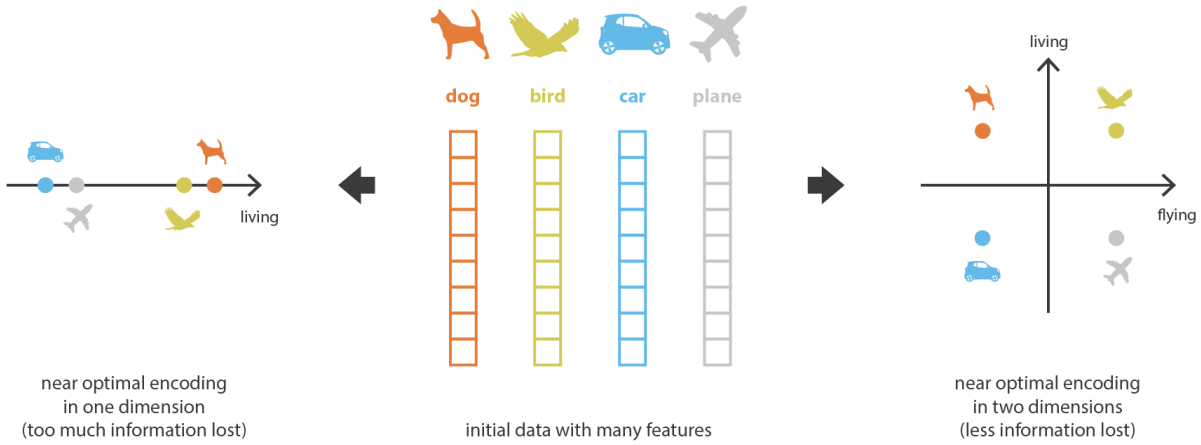
Zatem intuicyjnie ogólna architektura autokodera (enkoder + dekoder) tworzy wąskie gardło dla danych, które zapewnia, że tylko główna (najlepiej opisująca dane) ustrukturyzowana część informacji może przejść i być odtworzona. Wyszukiwanie enkodera i dekodera, które minimalizują błąd rekonstrukcji, odbywa się poprzez gradient descent na parametry tych sieci.

Co warto zapamiętać?

1. Redukcja wymiarów bez utraty rekonstrukcji często ma swoją cenę: brak możliwości interpretacji i wykorzystania struktury w utajonej przestrzeni (brak regularności).

2. Ponieważ przez większość czasu ostatecznym celem zmniejszenia wymiarów jest nie tylko zmniejszenie liczby wymiarów danych, ale zmniejszenie tej liczby wymiarów przy jednoczesnym zachowaniu większej części informacji o strukturze danych w zredukowanych reprezentacjach. Wymiar utajonej przestrzeni i „głębokość" autokodera muszą być dokładnie kontrolowane i regulowane w zależności od ostatecznego celu zmniejszenia wymiarów.

UWAGA: Zmniejszając liczbę wymiarów, chcemy zachować główną strukturę między danymi.

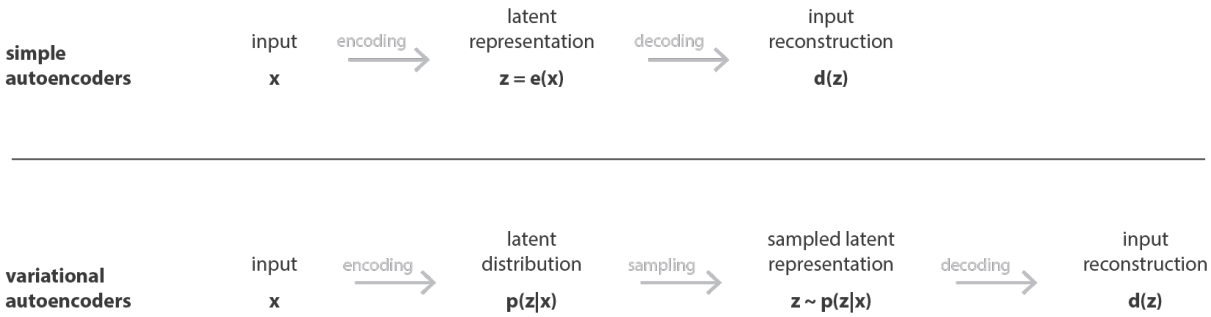


W tym momencie naturalne pytanie, które nasuwa się na myśl, brzmi „jaki jest związek między autoencoderami a generowaniem treści?". Rzeczywiście, po wyszkoleniu autoencodera mamy zarówno koder, jak i dekoder, ale nadal nie ma realnego sposobu na tworzenie nowych treści. I tu na ratunek przybywa VAE.

3 Autoenkoder wariacyjny (VAE)

Podobnie jak standardowy autoencoder, (VAE) autoencoder wariacyjny to architektura złożona zarówno z enkodera, jak i dekodera, która jest uczona w celu zminimalizowania błędu rekonstrukcji między zakodowanymi danymi a danymi początkowymi.

Aby jednak wprowadzić pewną regularyzację przestrzeni utajonej, przechodzimy do niewielkiej modyfikacji procesu kodowania-dekodowania: zamiast kodować dane wejściowe jako pojedynczy punkt, kodujemy je jako rozkład w przestrzeni utajonej.

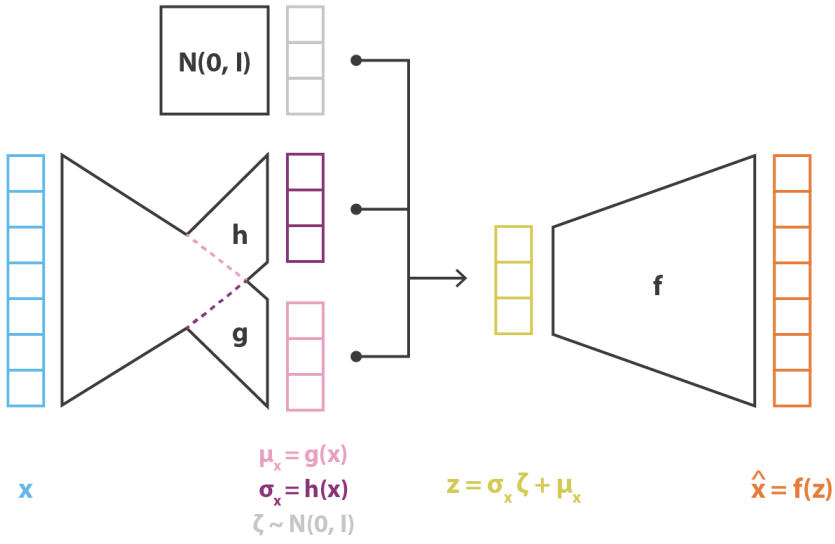


Rysunek 3: Różnica między autoenkoderem a autoenkoderem wariacyjnym.

Zatem model jest uczony w następujący sposób:

1. wejście jest zakodowane jako rozkład w utajonej przestrzeni
2. punkt z ukrytej przestrzeni jest generowany z tego rozkładu
3. wygenerowany punkt jest dekodowany i można obliczyć błąd rekonstrukcji
4. błąd rekonstrukcji jest rozpropagowany przez sieć.

W praktyce zakodowane rozkłady są wybierane tak, aby były normalne, dzięki czemu koder można wyszkolić do zwracania średniej i macierzy kowariancji opisujących te rozkłady.



Rysunek 4: Autoenkoder wariacyjny.

Zatem funkcja straty, która jest minimalizowana podczas szkolenia VAE, składa się z „terminu rekonstrukcji" (na końcowej warstwie), który ma tendencję do czynienia schematu kodowania-dekodowania tak wydajnym, jak to możliwe, i „terminu regularyzacji"

(na warstwie ukrytej), która ma tendencję do normalizowania ukrytej przestrzeni, ponieważ rozkłady zwracane przez koder są zbliżone do standardowych rozkładów normalnych. Ten termin regularyzacji jest wyrażony jako dywergencja Kulbacka-Leiblera między zwróconym rozkładem a standardowym rozkładem Gaussowskim.

4 β wariacyjny autoenkoder (β VAE)

Jeśli każda zmienna w wywnioskowanej utajonej reprezentacji z jest wrażliwa tylko na jeden pojedynczy czynnik generujący i względnie niezmienna dla innych czynników, powiemy, że ta reprezentacja jest rozplątana lub podzielona na czynniki. Jedną z korzyści, która często wiąże się z rozplątaną reprezentacją, jest dobra interpretowalność i łatwa generalizacja do różnych zadań.

Na przykład model wytrenowany na zdjęciach ludzkich twarzy może uchwycić delikatność, kolor skóry, kolor włosów, długość włosów, emocje, niezależnie od tego, czy nosi się okulary i wiele innych względnie niezależnych czynników w osobnych wymiarach. Taka rozplątana reprezentacja jest bardzo korzystna dla generowania obrazu twarzy.

Definicja: β VAE jest modyfikacją wariacyjnego autoencodera ze szczególnym naciskiem na odkrywanie rozplątanych czynników utajonych. Kierując się tym samym w VAE, chcemy zmaksymalizować prawdopodobieństwo wygenerowania prawdziwe dane, przy zachowaniu niewielkiej odległości między rzeczywistym a szacowanym rozkładem.

$$\text{loss}_{\text{VAE}} = C\,||\,x - \hat{x}\,||^2 - KL[\,N(\mu_x, \sigma_x), N(0, I)\,] = C\,||\,x - f(z)\,||^2 - KL[\,N(g(x), h(x)), N(0, I)\,]$$

$$\text{loss}_{\beta\,\text{VAE}} = C\,||\,x - \hat{x}\,||^2 - \beta\,KL[\,N(\mu_x, \sigma_x), N(0, I)\,] = C\,||\,x - f(z)\,||^2 - \beta\,KL[\,N(g(x), h(x)), N(0, I)\,]$$

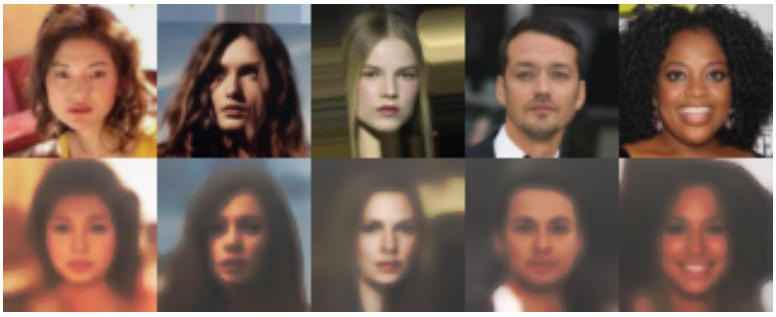
Rysunek 5: Funkcja straty w VAE i β VAE.

Gdy $\beta = 1$, jest to to samo co VAE. Gdy $\beta > 1$, stosuje silniejsze ograniczenie do ukrytego wąskiego gardła i ogranicza zdolność reprezentacji z. W przypadku niektórych warunkowo niezależnych czynników generujących ich rozplątywanie jest najbardziej wydajną reprezentacją. Dlatego wyższa β zachęca do bardziej wydajnego ukrytego kodowania i dodatkowo zachęca do rozplątania. Tymczasem wyższa wartość β może spowodować kompromis między jakością rekonstrukcji a stopniem rozplątywania.

UWAGA: Jednym z największych ograniczeń normalnego VAE jest to, że obrazy wyjściowe często nie są wystarczająco ostre.

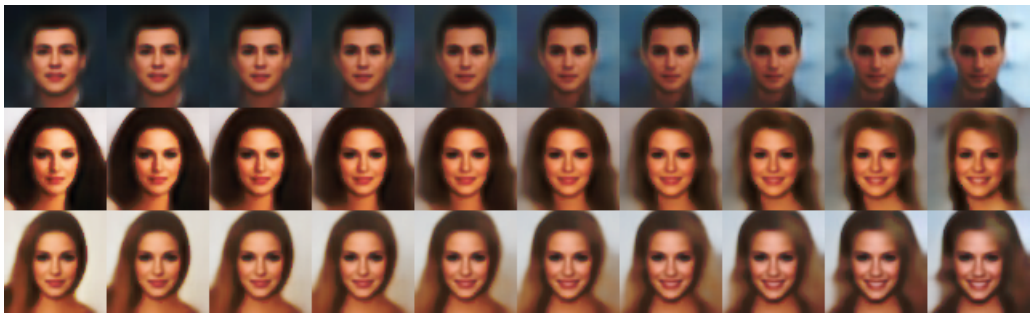
5 Przykłady działania β VAE

Używając β VAE możemy zrekonstruować zdjęcia.



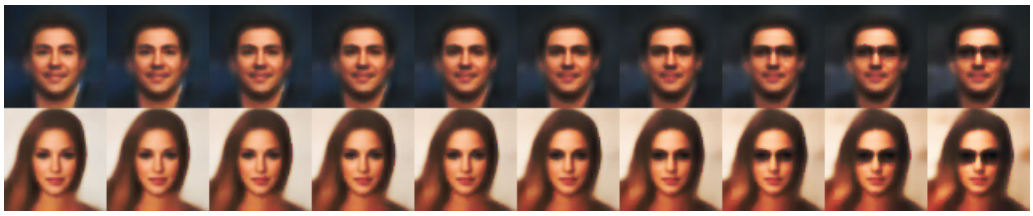
Rysunek 6: Oryginał na górze, rekonstrukcja na dole.

Możemy róznież zmieniać cechy wyglądu osoby na zdjeciu np. kolor, długość włosów, kolor oczu lub tło za osobą.



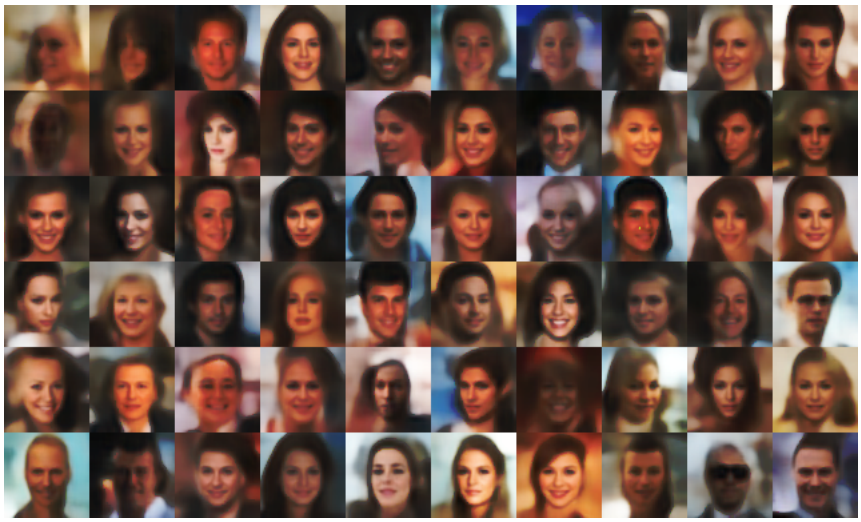
Rysunek 7: Interpolacja liniowa od z1 (skrajnie lewy) do z2 (skrajnie prawy).

Możemy nie tylko zmieniać cechy ale róznież je dodawać.



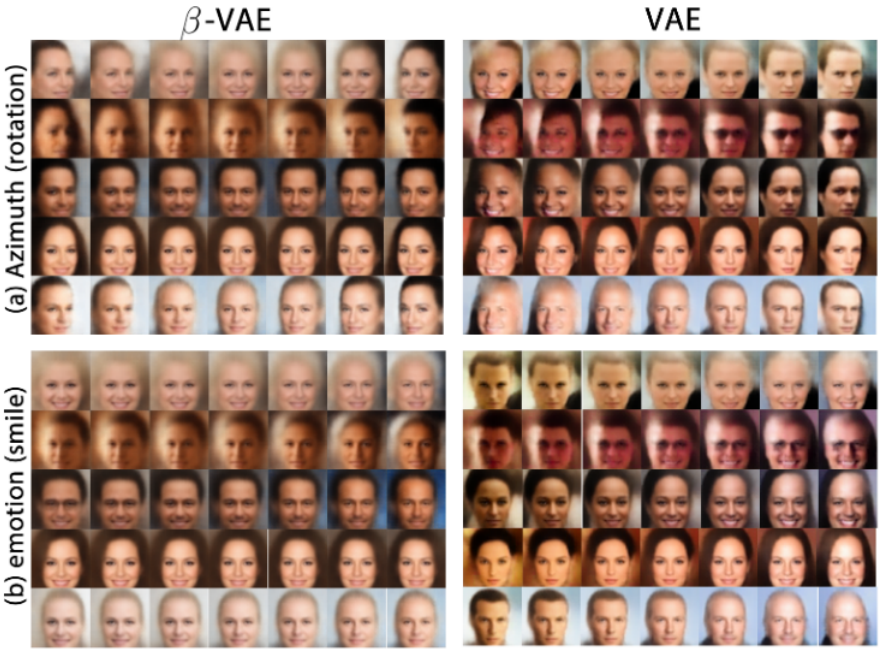
Rysunek 8: Oryginalne zdjęcia po skrajnie lewej oraz stopniowo dodajemy okulary, aż do stworzenia obrazu z osobą noszącą okulary po skrajnie prawej stronie.

β VAE dzięki kodowaniu danych na rozkład zamiast na pojedynczy punkt, umożliwia generowanie danych, które nie istnieją w rzeczywistości. Jest to bardzo pomocne przy obecnie obowiązującym RODO, kiedy chcemy np w tym projekcie pokazać przykłady zastosowania przedstawianych metod.



Rysunek 9: Wygenerowane obrazy z losowo wylosowanej z $\sim$ N(0,1).

Porównując VAE do β VAE możemy dojść do wniosku, że w VAE otrzymujemy ostrzejsze obrazy, ale β VAE działa lepiej z rozplątanymi ukrytymi czynnikami.



6 Bibliografia

1. Higgins I.,Matthey L., Pal A., Burgess C., Glorot X., Botvinick M., Mohamed S., Lerchner A.; "β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK"