# Master Analysis - Plots

*Marcin Kosinski*

*25.10.2015*

```
model_1_over_t$coefficients[c(98, 196, 294, 392, 490)] -> coeffs_1_overt
do.call(rbind, coeffs_1_overt) %>%
  as.data.frame %>%
  mutate(epoka = 1:5) %>%
  tidyr::gather(key=epoka) ->gat
names(gat)[2]<- "gen"
library(ggplot2)
gat %>%
  ggplot(aes(value)) + geom_histogram(fill=4, col = 1, binwidth =  0.01) +
  facet_grid(epoka~.) +
  theme_classic(base_size = 15) +
  ylab('liczność') + xlab('wartości współczynników modelu') +
  ggtitle('Histogramy wartości współczynników modelu po kolejnych epokach') +
  coord_cartesian(xlim = c(-0.5,0.5)) -> fig1


model_1_over_50sqrt_t$coefficients[c(98, 196, 294, 392, 490)] -> coeffs_over_50sqrt_t
do.call(rbind, coeffs_over_50sqrt_t) %>%
  as.data.frame %>%
  mutate(epoka = 1:5) %>%
  tidyr::gather(key=epoka) ->gat2
names(gat2)[2]<- "gen"
library(ggplot2)
gat2 %>%
  ggplot(aes(value)) + geom_histogram(fill=5, col = 1, binwidth =  0.01) +
  facet_grid(epoka~.) +
  theme_classic(base_size = 15) +
  ylab('liczność') + xlab('wartości współczynników modelu') +
  ggtitle('Histogramy wartości współczynników modelu po kolejnych epokach') +
  coord_cartesian(xlim = c(-0.25,0.25)) -> fig2

model_1_over_100sqrt_t$coefficients[c(98, 196, 294, 392, 490)] -> coeffs_over_100sqrt_t
do.call(rbind, coeffs_over_100sqrt_t) %>%
  as.data.frame %>%
  mutate(epoka = 1:5) %>%
  tidyr::gather(key=epoka) ->gat3
names(gat3)[2]<- "gen"
gat3 %>%
  ggplot(aes(value)) + geom_histogram(fill=6, col = 1, binwidth =  0.01) +
  facet_grid(epoka~.) +
  theme_classic(base_size = 15) +
  ylab('liczność') + xlab('wartości współczynników modelu') +
  ggtitle('Histogramy wartości współczynników modelu po kolejnych epokach') +
  coord_cartesian(xlim = c(-0.1,0.1)) -> fig3
```

```
gat %>%
  mutate(model = "model1") %>%
  bind_rows(gat2 %>%
              mutate(model = "model2")) %>%
  bind_rows(gat3 %>%
              mutate(model = "model3")) %>%
  ggplot(aes(value)) + geom_histogram(fill=4, col = 1, binwidth = 0.001) +
  facet_grid(epoka~model) +
  theme_classic(base_size = 15) +
  ylab('liczność') + xlab('wartość współczynnika modelu') +
  ggtitle('Histogramy wartości współczynników modelu po kolejnych epokach') -> fig4

do.call(rbind, coeffs_over_100sqrt_t) %>% t %>%  as.data.frame -> model3
do.call(rbind, coeffs_over_50sqrt_t) %>% t %>%  as.data.frame -> model2
do.call(rbind, coeffs_1_overt) %>% t %>%  as.data.frame -> model1
names(model3) <- c("epoka1", "epoka2","epoka3", "epoka4", "epoka5")
names(model2) <- c("epoka1", "epoka2","epoka3", "epoka4", "epoka5")
names(model1) <- c("epoka1", "epoka2","epoka3", "epoka4", "epoka5")

library(GGally)
ggpairs(  title = 'Różnice między współczynnikami w kolejnych epokach',
          model1,
          upper = "blank",
          lower = list(continuous = "points")
) -> fig5
do.call(rbind, model_1_over_t$coefficients) -> wspolczynniki
qplot(1:491,wspolczynniki[ ,which(colnames(wspolczynniki) == "BRAF")]) +
  xlab('Krok algorytmu') +
  ylab('Wartość współczynnika') +
  theme_minimal(base_size = 20) +
  ggtitle('Wartości współczynnika genu BRAF') -> fig8
qplot(1:491,wspolczynniki[ ,which(colnames(wspolczynniki) == "EGFR")]) +
  xlab('Krok algorytmu') +
  ylab('Wartość współczynnika') +
  theme_minimal(base_size = 20) +
  ggtitle('Wartości współczynnika genu EGFR') -> fig9
do.call(rbind,testCox) -> testCox_binded
do.call(rbind,trainCox) -> trainCox_binded
library(dplyr)
library(survMisc)
trainCox_binded %>%
  select(patient.vital_status, times, BRAF, EGFR) %>%
  survfit( Surv(times, patient.vital_status) ~ BRAF, data = .) %>%
  survMisc:::autoplot.survfit( titleSize=12, type="CI", palette = "Set1",
                               alpha = 0.7, survLineSize=2,
 pX = 0.3, sigP =3, title = "Przeżycie a mutacja genu BRAF" ) -> km_plot

xlims <- c(0,5000)
library(ggplot2)
km_plot[[1]] <- km_plot[[1]] +
  coord_cartesian(xlim = c(xlims[1],xlims[2])) +
  theme_light(base_size = 22)
km_plot[[2]] <- km_plot[[2]] +
```

```
   coord_cartesian(xlim = xlims)+
  theme_light(base_size = 22)
pdf("BRAF.pdf", width = 10, height = 8)
survMisc::autoplot(km_plot)
dev.off()

do.call(rbind,testCox) -> testCox_binded
do.call(rbind,model_1_over_t$coefficients)-> coeffM1
do.call(rbind,model_1_over_50sqrt_t$coefficients)-> coeffM2
do.call(rbind,model_1_over_100sqrt_t$coefficients)-> coeffM3
library(dplyr)
coxph_loglik <- function(beta, batchData ){
  batchData <- batchData %>% arrange(-times)
  scores <- apply(batchData[, -c(1092, 1093)], 1,
                  function(element) exp(element %*% beta) )
  distractor <- cumsum(log(scores))

  sum((as.vector(beta%*%t(batchData[, -c(1092, 1093)])) -
         distractor)*batchData[, "patient.vital_status"])

}

library(pbapply)
pbapply(coeffM1, MARGIN = 1, function(x){
  coxph_loglik(x, testCox_binded)
}) -> log_lik_M1
pbapply(coeffM2, MARGIN = 1, function(x){
  coxph_loglik(x, testCox_binded)
}) -> log_lik_M2
pbapply(coeffM3, MARGIN = 1, function(x){
  coxph_loglik(x, testCox_binded)
}) -> log_lik_M3

library(tidyr)
data.frame(model1 = log_lik_M1,
           model2 = log_lik_M2,
           model3 = log_lik_M3,
           krok = 1:491) %>%
  gather(krok) -> d2viz
names(d2viz)[2]<-"model"

library(ggplot2)
ggplot(d2viz, aes(krok, -value, group = model, col = model)) +
  geom_line() +scale_color_brewer(palette = "Set1") +
  coord_cartesian(ylim=c(0,4000)) +
  theme_light() +
  theme(legend.position = "top") +
  xlab("Krok algorytmu") +
  ylab("log-wiarogodność") +
  ggtitle("Wartości częściowej funkcji log-wiarogodności konstruowanej na zbiorze testowym
          wśród 3 badanych modeli, w kolejnych krokach algorytmu") -> fig10
```